Helsinki Studies in Education, number 78

Aleksi Kalenius

Are We There Yet?

Equality of learning opportunity in theory and practice

Doctoral dissertation, to be presented for public discussion with the permission of the Faculty of Educational Sciences of the University of Helsinki, in Auditorium Aurora 107, Siltavuorenpenger 3 A, on the 23rd of June, 2020 at 12 o'clock.

Helsinki 2020

Supervisors

Professor Patrik Scheinin, University of Helsinki

Professor Emeritus Jarkko Hautamäki, University of Helsinki

Custos

Professor Patrik Scheinin, University of Helsinki

Pre-examiners

Professor Benő Csapó, University of Szeged

Professor Emeritus Jan-Eric Gustafsson, University of Göteborg

Opponent

Professor Francesca Borgonovi, University College London

Cover picture

Aleksi Kalenius

The Faculty of Educational Sciences uses the Urkund system (plagiarism recognition) to examine all doctoral dissertations.

Yliopistopaino Unigrafia, Helsinki

ISSN 1798-8322

ISSN 2489-2297 (verkkojulkaisu)

ISBN 978-951-51-6155-0 (nid.)

ISBN 978-951-51-6156-7 (PDF)

University of Helsinki, Faculty of Educational Sciences

Helsinki Studies in Education, number 78

Aleksi Kalenius

Are We There Yet?

Equality of learning opportunity in theory and practice

Abstract

One of the fundamental questions of all societies is the possible trade-off between quality of outcomes and equality of opportunity.

The question of a possible trade-off between equality of opportunity and level of outcomes has a long history in the sociology of education, and it has been seen as a fundamental question of the field. Economics literature recognises that equality of opportunity is not, unlike equality of outcomes, in a trade-off with quality of outcomes. Rather, inequalities of opportunity generate inefficiency and are thus in a trade-off with the level of outcomes. Here, building on this difference, we make a distinction between *outcome trade-off* and *opportunity trade-off*.

Equality of opportunity is standardly measured indirectly, taking outcomes associated with circumstances to be due to circumstances and thus an indication of inequality of opportunity. This approach assumes that the circumstances of the individual are independent of the responsibility characteristics. The assumption flows from taking responsibility to be inseparable from individual freedom, whereby an individual should not be held responsible for the direct or indirect consequences of unchosen circumstances. When responsibility is directly linked to individual choice, we may also take any between-types differences in outcomes between groups facing different environments as evidence of the differential effects of those environments.

Equality of opportunity is a causal concept, decomposing the outcomes as due to just and unjust causes. Due to the complexity of social phenomena such as intergenerational mobility, *well-defined causality* is out of reach, yet the concept of opportunity requires counterfactuals and causality. We propose that, rather than abandoning the causal core of the concept of equality of opportunity, or high standards in identifying well-defined causality, complex circumstances should be understood as *ill-defined treatments*.

The area of education reveals interesting tensions because cognitive outcomes cannot be redistributed. It is usually taken to follow that differences in cognitive outcomes that are due to innate ability are not unjust, even though this means making individuals effectively responsible for something beyond their choice or control – their innate ability. This is what makes the opportunity trade-off possible. It arises because individuals in different circumstances may differ by innate ability.

To bridge the gap between responsibility linked to freedom and learning outcomes linked to innate ability, we propose a typology of responsibility characteristics to allow for a more explicit recognition of the reasons for holding individuals responsible for causes beyond their control. The core concept introduced is *natural responsibility characteristic*, referring to a cause that individuals are held responsible for, but beyond the influence of persons individually or

collectively. This makes natural responsibility characteristics historically contingent rules of regulation that depend on the level of technology, broadly construed.

The two hypotheses on the causes of the association between background circumstances and cognitive outcomes emphasise ability and opportunity. The *opportunity hypothesis* takes outcome differences associated with circumstances as being due to the opportunities that different groups face, while the *ability hypothesis* takes differences to arise from between-group differences in innate ability. Both innate ability and opportunity can be observed only indirectly, but the opportunity hypothesis predicts that the between-type difference is minimised with high outcomes, while the ability hypothesis leads us to expect that equality of opportunity and the level of outcomes form a trade-off. The two perspectives can be combined in an *ability-opportunity curve*, the peak of which depends on the contribution of innate ability differences to between-group differences in outcomes.

The causal chain between genetics and cognitive outcomes is insufficiently understood to provide clarity to the role of genes in explaining between-group differences in cognitive outcomes. While all researched human traits are heritable and we find an association between genetics and outcomes, unobservable associations between environment and genetics currently prevent well-defined identification of the causal contribution of genetics to between-group differences.

The empirical sections of the work approach the question of equality of learning opportunity and the level of learning outcomes using IALS and PIAAC, two studies of adult skills. The ability-opportunity curve is approached, observing the association between equality of opportunity and the level of learning outcomes for different parental education types.

The results indicate a positive association between equality of learning opportunity and the level of adult skills for all parental education types, matching the prediction of the opportunity hypothesis. While high proficiency can be achieved with high levels of equality of opportunity, the two usually go together. The positive association between equality of opportunity and the level of learning outcomes is stronger for lower parental education types and on longer time frames. The longer-term equality of opportunity is a strong predictor of the level of learning outcomes even in the short term. Equality of outcomes, here captured by income inequality, explains much of the same variation as equality of opportunity, supporting the opportunity-equalising effects of income equality attested to by the *Great Gatsby curve*.

Longer-term equality indicators also predict learning outcomes of the high parental education type over time to the same degree as contemporary equality indicators, implying that equality indicators may provide advance warning of the direction of the evolution of learning outcomes. The results imply that developed countries still find themselves on the upward-sloping section of the ability-opportunity curve.

In the analysis of possible traces of education reforms in between-cohort differences in equality of opportunity, IALS and PIAAC paint a largely concordant picture of the evolution of proficiency across cohorts. The pattern applies to the youngest groups of IALS, implying that cohort discontinuities in adulthood may be associated with differences in educational experiences in youth and with policy reforms. Comparing the timing of the observed discontinuities with the timing of education policy reforms, we identify 29 discontinuities that can be associated with notable changes in education policy. Reforms towards longer education and more comprehensive schooling seem mainly to be associated with improved equality of learning opportunity, while in some countries similar reforms in the 1990s seem to be associated with declining equality of opportunity.

A brief, slightly more in-depth, look at Finland reveals a large increase in inequality in the 1990s, which broadly matches other data sources. The observation is of interest for several reasons. It would, if the results of our analyses are anything to go by, predict a later decline in learning outcomes, something that Finland has famously observed over the last 13-25 years. In addition, many of the internationally-known characteristics of the Finnish education system have been linked to the reforms of the 1990s and those reforms have been used to explain the rise of Finland to international fame and to the position of a global benchmark. The timing of the increase in inequality matches these reforms. The results suggest that the causes of Finnish success might have been misidentified and that we may have, both nationally and internationally, learnt the wrong lessons from Finland.

Keywords: equality of opportunity, learning outcomes, large-scale survey, adult skills, education policy

Foreword

This was not the way the work was supposed to turn out. The idea was to go straightforwardly to the policy-oriented question of the trade-off between equality of opportunity and the level of learning outcomes. To answer the question: do we have to choose between equality and quality?

The purpose was to apply standard methods for measuring equality of opportunity to the measurement of equality of learning opportunity and use the results as a starting point for various excursions into the relationship between equality of learning opportunity and policy as well as the trade-off between equality and the level of learning outcomes. Something quick, something straightforward. Something that would fit nicely in a few small articles.

We aim for the good, but things turn out as they always do.

Equality of opportunity is a multifaceted concept, raising the question of where we should, as a matter of justice, draw the line between individual responsibility and circumstances that individuals are not held responsible for. Different choices on where to draw the line lead to substantially different conceptions of justice – to different normative commitments and to very different identification strategies – that only share the commitment to allow just individual differences due to just causes, and to identify and remove unjust causes of differences.

Opportunity is a hypothetical – a concept to describe the counterfactual that cannot be directly observed. We may, of course, observe various factors that contribute to opportunity – free schooling contributes to opportunities to learn – but with the multitude of causes that contribute to or deduct from the possibility to learn, such an approach either leads to an impossibly extensive list of contributors to opportunity or to a narrow conception of opportunity. A non-procrustean conception of opportunity is not about what is, but about what could be and what could have been. But building the counterfactual requires a causal model that can predict the outcomes given different causes. For questions of intergenerational mobility and social equality, the details of such a causal model – why and how individuals end up in positions and with outcomes that seem to be associated with their background – are far beyond the reach of current research.

The focus on learning also invites in all the complexity of the question of innate ability. A focus on a redistributable outcome like income, or aggregate welfare, where shortcomings in one respect can be compensated by a better position in another, makes it possible to achieve equality through social choice. That allows a full range of normative positions, from holding individuals responsible for their innate ability and treating any proceedings from that ability as just – to concluding that innate ability is unearned and thus an unjust cause for differences in outcomes. In the domain of learning, not all choices are available, because learning outcomes are inalienable and cannot be redistributed. This seems to leave two options available: holding individuals responsible for things beyond their control or considering differences in learning outcomes due to innate ability to be unjust. As both options are very problematic, it was necessary to find a normative treatment that would recognise the necessary role of innate ability, while not severing the link between responsibility and the freedom to act otherwise.

The question of the trade-off between equality of opportunity and level of learning outcomes has been a fundamental question of policy and research. Not only because of the theoretical, but also because of the empirical difficulties. When opportunities are explored on the level of groups, taking group membership as a proxy for the treatment of an individual, it is

unclear what the treatment effectively consists of. Over the years, opinions have ranged widely, with hereditarians taking class, race and gender differences as being due to innate differences between the groups, while environmentalist approaches sought causes from the different social positions and circumstances of the groups. These discussions go back to the ancients, and to the fundamental question of the basis of social positions and hierarchies.

Thus, a simple question brings up numerous complex, complicated and unresolved normative and empirical questions that affect the interpretation of any empirical results. Moreover, the complexities are intertwined, so that the question of unobservables in measuring opportunity is linked to the heritability of cognitive traits in several ways. As many of the conundrums cannot be solved or bypassed, they can be treated either implicitly or explicitly. In an implicit treatment, it would be possible to present the choices and the process in technical terms, without trying to normatively justify the chosen normative approach. I have made a choice for the explicit, rather than the implicit, making an effort to make many of my choices explicit and to justify them. This approach has resulted in an extensive exploration of the marshy grounds between normative and empirical enquiry. While many of the key questions must remain without final answers, I have felt it to be worthwhile, even necessary, to find conceptual structures to make sense of the many dimensions that the question brings up.

The search has led me to explore various discrete domains of research literature. The normative dimensions of the question – equality, freedom and justice – have led me back and forth through philosophical and juridical literature. Empirical measurement of equality of opportunity has led to excursions in the economics literature on measuring equality of opportunity, but also to the sociological literature on intergenerational mobility and to general questions of causal inference. The challenge of the role of innate ability in explaining differences between groups has led to psychometric literature on intelligence and behavioural genetics. In all fields, questions have come up that have forced me to turn around and slog back to other areas of literature to make sense of how the different pieces of the puzzle fit together.

The nature of the interwoven questions, and my unwillingness to leave significant normative or other commitments unaddressed, has led me to dabble in questions that fall into many different fields of inquiry. Rather than hiding substantive choices behind technical language, or technical choices behind noble intentions, I have tried to cut through the complexity by presenting interpretational schemas that have helped me to make sense of the complexities involved. While all this started with the empirics, it was in the end at least as much about the theoretical background as it was about the empirical findings.

As with any work, there is a persistent temptation to explore further. To clarify and condense, to explicate and expand, to address questions left unexplored. The work opens up in many directions and explicitly addressing all the relevant questions would be impossible, as each level brings up new fundamental questions to be addressed. Now, many of the more immediately relevant questions have been addressed, perhaps even some questions that are more peripheral. Thus, it is time to take stock of what has been achieved, before continuing further. For now, many of the different conceptual and empirical pieces fit – not in any arrangement that could be considered final, but well enough to take a snapshot of the current arrangement before continuing with the exercise of finding more and more pieces to the puzzle, understanding better the pieces already in it, and moving the pieces back and forth to try to put the picture into sharper focus.

My gratitude goes to my friends and colleagues, who have helped and encouraged me on this project over the years. I would particularly like to thank my advisors, Patrik Scheinin and

Jarkko Hautamäki, and the pre-examiners Benő Csapó and Jan-Eric Gustafsson for their insightful and encouraging reviews of the work. I'm grateful to Francesca Borgonovi for kindly agreeing to act as my opponent at the public examination of my thesis. I would like to thank Antero Malin, Venla Bernelius, Hannu Karhunen, Mari-Pauliina Vainikainen, Sirkku Kupiainen, Antti Latvala, Lars Nerdrum and Daniele Chechhi, for their critical, constructive and helpful comments on this monograph or the initially separate pieces of work that it evolved from. I'd also like to extend my gratitude to the Faculty of Educational Sciences that kindly provided me the opportunity to take a leave of absence from other professional endeavours and devote half a year to the finalisation of the manuscript. Finally, my thanks go to Kenneth Quek for revising and polishing the language of the finalised manuscript.

Most importantly, I am grateful to my wife Saija and my children Tyyra, Katla and Enki. My gratitude is not eternal – nothing ever is – but it is as eternal as it is in my power to feel. I'm grateful for all the patience they have shown and support they have given me in the course of this work, especially when the finalisation phase took me to live in another city in another country. I'm grateful for all the patience and support that they continue to show and give to me. And for making me blissfully happy to be alive, here and now, from opening my eyes every morning to closing them in the evening. Nothing else is quite as real, quite as true, or quite as just. Very special thanks go to my wife, who has had the quite questionable pleasure of reading the text in several phases, when I have needed another set of eyes to spot even glaring mistakes in a text that I have become blind to myself. The task does not reward itself, and I have been lucky to have her help.

All mistakes, faults and shortcomings in the following are mine and mine alone.

Aleksi Kalenius
In Helsinki on May 12th 2020

Table of contents

Introduction

Is there a trade-off between equality of opportunity and the level of learning outcomes? Do we have to sacrifice one to gain the other? The question is simple enough – to answer it is anything but. This work attempts to shed some light on how we should understand the relationship between equality of learning opportunity and the level of learning outcomes both theoretically and empirically, through the keyhole of adult skills surveys.

The theoretical part of the work, consisting of sections I to III, explores how the concept of equality of opportunity is to be understood in the context of outcomes that are partially determined by biological factors beyond human control, how this understanding is related to what is just or unjust, and to the possible trade-off between equality and outcomes. The empirical part of the work uses large-scale surveys of adult skills to analyse the observed relationship between equality of opportunity and the level of learning outcomes and the changes of equality of opportunity associated with changes in education policy.

Equality of learning opportunity

The theoretical part engages in philosophical and methodological discussion on the concept of equality of opportunity in the context of learning outcomes. The treatment is quite extended because many of the pertinent questions still require theoretical clarification. The role of innate ability in influencing learning outcomes brings many interesting questions to the fore, when we try to apply the concept of equality of opportunity to the area of learning outcomes.

Standard empirical methodologies of identifying equality of opportunity proceed in several steps.[1] They first identify the causal model that determines the outcome distribution and then measure the counterfactual distribution of outcomes that the model produces when the effect of fair causes has been eliminated, i.e. when only unfair causes affect the distribution of outcomes. They distinguish outcomes between those unfair inequalities due to circumstances beyond individual control and fair inequalities due to individual choice and effort. A crucial assumption in this decomposition is of independence of circumstances and morally accountable effort. This assumption is reasonable when individuals are to be held responsible for characteristics under their control, because then all association of observed effort and circumstances may be taken as being due to the effects of circumstances on individual effort through incentives.

The matter becomes complicated when causes beyond individual control are taken to be just causes of the distribution of outcomes. Differences in innate ability are such causes, as individuals cannot choose their innate abilities, but these are often taken to be just – or at least not unjust – causes for learning outcomes. Thus, while it is rarely explicitly recognised in the literature, standard methods of measuring inequality of opportunity do not apply to the area of learning without additional theoretical commitments or empirical assumptions. One fundamental assumption pertains to the justification for holding individuals responsible for causes beyond their control, another to the distribution of innate ability between social groups.

[1] The introduction is free of references to the literature. It provides an overview of the entire work and replication of the entire reference apparatus would make it unbearably heavy. References are thus to be found in the main body of the text.

We present a case for considering innate ability to be a natural responsibility characteristic. By this concept we refer to a provisionally just cause of the distribution of learning outcomes. It is just by virtue of being unavoidable, as innate ability is inalienable and non-malleable, but only just as long and as far as it is a question of natural necessity. To the degree that technological progress makes the association between innate ability and learning outcomes more malleable, the moral arbitrariness of innate ability makes it a circumstance that should be equalised. The concept of a natural responsibility characteristic acts as a bridge between Rawlsian fair equality of opportunity, which has individuals occupying positions that correspond to their innate abilities, and the Rawlsian position of treating innate ability as a natural resource and thus a common asset. Any unavoidable link between innate ability and learning outcomes makes equalisation of learning outcomes possible only by levelling down so that greater equality would entail lowering the outcomes of the talented. To avoid levelling down, we may consider innate ability as a just cause of learning outcomes and of allocation of positions in the social division of labour. However, because the link between innate ability and welfare is not given by nature but determined by social arrangements, innate ability cannot be considered a just cause for differences in welfare between individuals.

This position for consideration of the role of innate ability as a responsibility characteristic and a circumstance is grounded in the distinction between fact-insensitive and fact-sensitive principles of justice and the corresponding distinction between fundamental principles of justice and rules of regulation as the practical application of those principles in a given situation. While it is unjust that individuals are responsible for causes beyond their control, it is unjust as a fact-independent principle of justice. Differences in learning outcomes due to differences in innate ability are just, as a matter of rules of regulation that take into account the natural constraints that might affect the full realisation of principles of justice, because innate ability cannot be redistributed.

Recognition of this dual nature of innate ability in treatment would allow a more nuanced and explicit treatment of the core normative and empirical questions of equality of opportunity. It clearly delineates the conditions for treating innate ability as a responsibility characteristic. It also has broader implications for theories of justice because it brings into the open the critical role of the location of the boundary between natural and social determinants of outcomes in determining the justice of causes. It helps to recognise why some causes that are just for the distribution of outcomes or opportunity in one historical situation might not be just in another, because what can be socially determined by morally competent agents depends on natural laws and the level of technology to affect them. This highlights the limited nature of any conception of justice that would wish to make transhistorical and universal principles of justice out of rules of regulation that have had to make do with the crooked timber available to construct the best version of justice possible in one historical situation, and recognises that, whilst our technical skills of forestry and carpentry improve, we can construct better forms of justice and do not have to accept injustice simply because justice was previously unachievable.

The fact that measuring inequalities of opportunity with counterfactuals – by the distribution of what could have been – requires identification of a causal model of determination of outcomes. It is generally identified through decomposition of the variation of outcomes, and this has led to some confusion in the literature between equality of opportunity and equality of outcomes. This confusion is partially linked to the fact that equality of opportunity literature insufficiently recognises the role of causal inference in identifying inequality of opportunity. The literature refers to measuring inequality of opportunity through the covariation of outcomes and circumstances, taking variation of

outcomes with circumstances to be unjust, because circumstances are unchosen. The crucial choice is, however, taking the observed associations between circumstances and outcomes to be causal factors in a causal model. The variation *with* circumstances is taken to be variation *due to* circumstances. Because equality of opportunity measures inequality due to unjust causes, it needs to assess what variation of outcomes is due to unjust causes. Thus, the circumstances of one's background become interpreted as a treatment, and the question of equality of opportunity becomes a question of equality of different treatments.

We argue that the cautious attitude towards the causal interpretation of identifying inequality of opportunity is due to general avoidance of making causal interpretations regarding associations in observational data. This caution is well-grounded in the fact that causal interpretation of well-defined interventions requires a setting that eliminates the effects of unobservables.

But inequality of opportunity is a meaningful concept only with a causal interpretation. Rather than abandoning a causal interpretation of the project of measuring inequality of opportunity, a strategy should be adopted that explicitly recognises that the process of causal inference consists of a pre-identified causal model, which is then specified by observing the strengths and effects of the associations in the data. This application of causal calculus would allow the field to further develop theoretical grounding of the measures of inequality of opportunity.

Where we remain in ignorance is the exact nature of the treatment in question. Upstream causes of outcomes, such as socioeconomic background, are understandable as ill-defined treatments. Such causes are largely beyond experimental, and even quasi-experimental, designs that would be able to eliminate the confounding effects of unobserved variables. Explicit recognition of ill-defined treatments would, however, allow us to recognise that the so-called achievement gap, the range of outcomes of the best-off and worst-off types, is equivalent to the average treatment effect when the defined type represents a treatment. An ill-defined one, but a treatment nonetheless.

The treatment of these questions occupies the first five chapters, forming the first two sections, of the current work.

First section – identifying opportunity

The first section presents the questions of normative and empirical identification of opportunity. These form the bedrock on which all treatments of equality of opportunity are built. The section is divided into two chapters, the first of which focuses on equality of opportunity as a normative ideal while the second focuses on empirical measurement.

Presentation of equality of opportunity as a normative ideal in **Chapter 1** provides an overview of the philosophical discussion on whether equality is a normative ideal. On Senian lines, we consider all discussions on the justice of equality to be discussions on the right equalisandum as a matter of justice. All conceptions of justice tend to agree that individuals are to be made equal in something, even if that something would be a very minimal set of process rights. Correspondingly, all conceptions of justice hold individuals responsible for some but not all causes of their outcomes. After a quick overview of the ideal of equality as the silence of reasons, the chapter turns to the concept of responsibility. Responsibility is usually justified by individual freedom, which is taken to imply not only the opportunity to make one's own choices but also to bear their consequences and receive praise or blame for them. This implies a reward principle whereby it is just that individuals receive rewards relative to their merits.

The combination of equality and responsibility in responsibility sensitive egalitarianism or equality of opportunity is presented through Temkin's canonical egalitarianism and Arneson's moral meritocracy, which combines the principle of vertical equity – that it is just for unequally deserving individuals to have different outcomes – with horizontal equity – the justice of treating equals equally. Moral meritocracy amounts to a minimalist version of egalitarianism, whereby justice prevails when differences in holdings correspond to differences in responsibility characteristics and are unaffected by differences in circumstances. This is the basic concept of equality of opportunity used in the later analysis.

In **Chapter 2,** we turn to the empirical measurement of equality of opportunity. The first of the two steps needed to identify equality of opportunity is the identification of a causal model that determines the outcomes of interest. Because equality of opportunity holds individuals responsible for only some of the causes of their outcomes, the causal model is needed to produce and measure, in the second step, a counterfactual distribution that reflects only the effects of unjust causes.

The causal model is produced by combining a previously identified causal structure and identifying the associations of interest from the data. Thus, the observed association between background characteristics and outcomes is interpreted as causal, and the variation *with* background characteristics as variation *due to* background characteristics.

The presentation focuses on the between-types inequality of opportunity methodologies, in which the causal model is identified by the statistical association of environmental indicators with outcomes. The population is divided into types sharing the same circumstances, such as ethnic group, sex or parental education. Differences within types are taken to be due to effort and variation between types to be due to circumstances. This identification of statistical dependency as the causal model allows the creation of a counterfactual distribution, which removes any within-type variation of outcomes. The process leaves only the variation of outcomes with circumstances, which is taken to represent the causal effect of those circumstances on outcomes and act as a measure of inequality of opportunity. The resulting counterfactual distribution can then be measured using any relevant measure of distribution.

In this approach, circumstances and normatively pertinent effort are assumed to be independent, in the sense that any indirect association of background and outcomes through effort is interpreted as arising from behavioural responses to circumstances.

The second section – Innate ability and responsibility without freedom

In the second section, we turn to the normative challenges of only holding individuals responsible for causes within their control when it comes to equality of learning opportunity. We know that individuals differ in innate ability, and that innate ability is by definition beyond individual control. At the same time, it would be counterintuitive to interpret that differences in learning outcomes due to differences in innate ability represent morally objectionable inequalities of opportunity and to demand that outcomes be equalised on a level achievable to everyone.

Chapter 4 delves into the basics of normative judgement. Any normative judgement arises from raw facts of the state of the world and of normative principles. As normativity is independent of how things are, fundamental normative truths are fact-insensitive, i.e. not sensitive to any raw facts of the world. These principles are then applied when we normatively assess any state or process in

the world. This gives rise to two different levels of normative principles of justice. The fact-insensitive *fundamental principles of justice* may be impossible to fulfil. *Rules of regulation*, in turn, are principles of justice that are just as applied to a given society, partially because they take into account the prevailing raw facts of that society. Some things may be taken as just, or not unjust, by a society that is unable to change this state of affairs and thus has to accept it, irrespective of any purely normative considerations.

Chapter 5 addresses directly the question of taking innate ability as a responsibility characteristic. There is widespread agreement that innate ability is morally arbitrary in the sense of being outside individual control, but some differences in approach as to what this entails. The chapter presents a proposed typology of different types of causes and how they are morally valued. The typology classifies causes into the natural (beyond human control), the social (in human control, but beyond individual control) and individual causes (individual choice and action). This typology allows for a more explicit recognition of the fact that we may, and often do, hold individuals responsible for causes beyond their control. The chapter also tries to identify principles and criteria by which we may do so in terms of rules of regulation. The core concept introduced is the natural responsibility characteristic, referring to a cause that individuals are held responsible for, but beyond the influence of persons individually or collectively. It is a principle of rules of regulation, being justified by the fact that it is unavoidable.

The typology is then briefly used to highlight that equality of opportunity as a theory is characterised by the conditions under which individuals are held responsible for something. This allows us to identify the critical differences between substantive equality of opportunity and procedural justice. Secondarily, the typology draws attention to natural responsibility characteristics being such due to the lack of counterfactuals, by virtue of being unavoidable. This makes natural responsibility characteristics historically contingent rules of regulation that depend on where the level of technology – i.e. the tools at our disposal – sets the limits on affecting the causes of outcomes. In a brave new world where innate ability was malleable, the rule of regulation concerning just distribution of innate ability might well be different.

The chapter closes by presenting how the principles of holding individuals responsible dovetail with the empirical measures of equality of opportunity. Empirically, most treatments of equality of opportunity include innate ability in the potential causes of confounding, or hold individuals responsible for some of their innate ability but not all. This highlights the interplay of normative and empirical challenges. The unobservability of innate ability makes standard empirical measures based on the covariation of circumstances and outcomes challenging, because selection by ability may not be controlled for. The same empirical strategy thus identifies – depending on the prevailing association of ability and circumstances – widely differing normative constructs. Under the same strategy, individuals are held responsible for very different things if circumstances are associated with the background than if they are not.

This section further specifies the adopted conception of equality of opportunity so that each individual is held responsible, as a rule of regulation, for their innate ability, because it is a natural responsibility characteristic.

Trade-offs and natural aristocracies

The second leg of the theoretical journey undertaken in the first part of the book covers the territory of the social trade-offs related to pursuing the twin goals of quality and equality of opportunity. Any possible trade-off between the two is a fundamental question of education policy, and most areas of inquiry into learning, because it would entail that society must make a valuation as to what degree it is willing to sacrifice excellence or quality of outcomes to achieve equality.

While the fundamental nature of this trade-off is recognised in the literature, it is less well recognised that any possible trade-off between equality of opportunity, understood as the independence of outcomes from social backgrounds such as race or social class, can only arise if the groups or types in question are systematically selected by innate ability. Indeed, the question of the equity-efficiency trade-off in most education policy discussion is not a discussion of outcome trade-off where equalisation of outcomes leads to levelling down, but a discussion on innate ability differences between historically privileged and underprivileged socioeconomic or racial groups.

Thus, the empirical challenges of identifying innate ability and opportunity – both of them counterfactual concepts that cannot be observed directly – are intimately tied to very concrete questions on whether we can expect to achieve better outcomes when pursuing equality of opportunity or only a gradual levelling-down. Or, more generally, whether current social hierarchies are a reflection of underlying natural hierarchies or not. After all, any implications of a measure of inequality of opportunity that takes background as treatment are substantially affected if the treatment is genetic rather than environmental. It is a different question if we have to be able to manipulate base pairs than if it is enough to manipulate our surroundings.

Treating the equity-efficiency trade-off in education in this context reveals that it is intimately tied to the age-old discussion on social and natural aristocracies. The hereditarian ability hypothesis has explained social outcomes with innate differences between groups, while environmentalists have emphasised environmental differences. Hereditarians have taken between-group differences to be determined by the same causes that determine within-group differences and have explicitly or implicitly assumed that heredity describes the determinative causal effect of genes on outcomes. Thus, as intelligence has been interpreted to be as much as 80 % heritable, the group differences in outcomes have been interpreted to be primarily determined by genes and beyond remedy by environmental intervention. Environmentalists have emphasised that groups differ by their environments as well and that heredity does not measure the malleability of a trait by environmental manipulation.

The hereditarian/environmentalist discussion on the origins of between-group differences in cognitive outcomes is linked to the discussion on the malleability of cognitive characteristics, but should not be confused with it. Any theory of primarily environmental determination of cognitive outcomes requires that cognitive performance is malleable by environmental interventions, and any theory of hereditarian bent requires that individuals inherit genes that exercise a non-malleable effect on their cognitive outcomes. While heredity is not a measure of the fixity of a trait, it often acquires significance in explanations of between-group differences when it is explicitly or implicitly interpreted as one. The identification of heredity with causal and fixed effects on phenotype rests on an interpretation that, while phenotypes emerge from a complex interplay of genotype with the environment, environmental effects are – at least from the perspective of intentional policy interventions – essentially random. This interpretation of the nature of the interplay between genetics and environ-

ment is the essential bridge from the statistical concept of heredity – the association of genetic variation with phenotypic variation – and malleability. It would mean that, even if the phenotype is heavily influenced by the environment, these environmental influences amount to Brownian motion without a direction, leaving genetics to give that direction. The randomness of these influences leaves them without a handle. As we do not have a handle on them, we may not use them to influence cognitive outcomes in turn. But we do not know if the bridge is there to be traversed or not.

The question of ability or opportunity determining group differences in outcomes does not turn simply on humans being or not being blank slates, written on by the environment. It is generally agreed that humans are indeed not blank slates. It also requires that groups differ (markedly) in their inherited innate ability, that some socioeconomic or ethnic groups have different non-erasable markings on their slates than others. This is where the current understanding of behavioural genetics still leaves us very much in the dark, which may explain why many of the tensions persist that were already present well before the current epoch of molecular behavioural genetics.

First, we are still far from understanding how genetics influences complex behavioural traits. Or any complex traits, for that matter. We have known since the era of twin studies that all traits are heritable, most of them to a considerable degree. But heredity is not a measure of fixity of traits, and environmental changes may have very significant effects on highly heritable traits – height being the most prominent and well-known example.

The lack of understanding of the causal chain between a gene and a complex trait means that we have limited understanding of the importance of the constraints posed by the genome on the possible range of outcomes relative to the effects of the environment. It has been argued that we may not even analytically differentiate between genetic potential and environmental effects on the realisation of that potential, as potential is deeply environment-specific and genomes providing an advantage in some conditions might not provide one in others. Moreover, we know that genetic influences are realised in an interplay with the environment, where gene-environment interaction works both ways. Not only does the environment affect the manifestation of genetic traits, but genes also affect environments through parents and self-selection into and active construction of environments. It is very uncertain how much of an interventional, policy-usable handle on such environmental effects we have. We currently do not have any analytical measure that would measure biological fixity of traits without further assumptions.

While some environmental effects are essentially random and thus beyond influence, we also have evidence that systematic system-level environmental interventions may have a very significant effect on, e.g., intelligence. The best available quasi-experimental evidence suggests that one year of schooling provides 1-5 points (.7-.33 SD) of increase in intelligence, which is far from insignificant when the populations of developed countries approach an average of 15 years of initial formal education.

Because the possibility of experimental designs is minimal, even non-existent, in cognitive genetics, one approach to the interpretation of the origins of between-group differences in innate ability has been to follow change over time. There, *in lieu* of assuming that heredity in one environment is effectively indicative of a genetic origin of differences between groups, the approach is to see how different groups develop relative to each other. Such observational work has different uncertainties than extrapolating within-group interpretations of genetic influences to between-group interpretations. It mainly consists of trying to assess, using different methods of triangulating evidence, whether group differences are on a track that may eventually lead to their elimination or not.

The third section – Explaining society and history

In the third section, the question of equality of opportunity and confounding of circumstances and innate ability is treated as a systemic social and policy question. The literature acknowledges that the equality-efficiency trade-off arises from the behavioural responses of actors, as equalisation of outcomes reduces incentives to bring about those outcomes. The trade-off between equality of opportunity and level outcomes emerges, in contrast, if types also represent selection by innate ability. The literature contains a wide variety of positions on the relative contributions of innate and environmental causes to life outcomes. The environmental approach takes environments to be causal, while the hereditarian approach takes association of circumstances and outcomes to be an artefact of selection by ability.

Chapter 6 starts off by presenting the equality-efficiency trade-off, in which equality of outcomes is associated with inefficiency. It is noted that inequality of opportunity, as inequality of circumstances, produces inefficiency, as it leads to misallocation of resources; thus, equality of opportunity is not in a trade-off with outcomes. Thus, while the question of a potential trade-off between socioeconomic equality of opportunity and the level of outcomes has been one of the fundamental questions of education policy and sociology, it is far from clear that we should expect a trade-off between the two. The inclusion of innate ability changes this picture, as a trade-off between equality of opportunity and level of outcomes may arise if the relevant groups differ in innate ability. Here we propose making a distinction between outcomes, opportunity and evolutionary trade-offs. The first is the traditional equality-efficiency trade-off, the second arises from innate differences between the relevant groups, and the third represents an evolutionary balance where the first two balance each other out.

After noting that the emergence of opportunity trade-off is dependent on between-group differences in innate ability, ability and opportunity hypotheses are presented as two classical hypotheses on the causes of differential attainment of different groups. The opportunity hypothesis explains differential outcomes through differential access to opportunity, in the same way as standard methods of measuring equality of opportunity. The ability hypothesis, in turn, explains intergenerational transmission of status through inheritance of innate and permanent characteristics. The chapter introduces the Ability-Opportunity Curve as a heuristic device for representing hypotheses on the contributions of innate characteristics and circumstances to individual or social outcomes.

Once the hypotheses have been presented, **Chapter 7** reviews the current state of evidence on the biological basis of differences in social, primarily educational, outcomes. The heritability of all behavioural traits is noted, as is the evidence that intelligence and other indicators of cognitive outcomes or performance are malleable and responsive to intervention. The main focus of the chapter lies in identifying evidence of how much observed between-group differences could be considered to be innate and not responsive to environmental interventions. It is concluded that the causal paths linking genetics and cognitive outcomes are still insufficiently known either to explain group differences in outcomes with genetics or to reject the possibility of such an explanation.

Empirical results

After the first three sections have paved the way by exploring the concept of equality of opportunity, justifying treating innate ability as a responsibility characteristic, and reviewing the current state of

evidence on the genetic basis of between-group differences in innate ability, it is time to turn to the empirical analyses on the opportunity trade-off and the effect of education policy on equality of opportunity.

The data and methodology are presented in **Section IV**, with research questions as well as the data and the measures used, in **chapter 8** and the analyses conducted in **chapter 9**. Empirical results are presented in **Section V**, with **chapter 10** presenting the results of analyses using equality of opportunity to predict proficiency, and **chapter 11** constructing a model to predict level of learning outcomes with equality indicators. Finally, **chapter 12** presents results on the relationship between equality of opportunity differences between cohorts and education policy changes over time.

The first phase of the analysis explores the potential opportunity trade-off for different parental education groups. The analyses are not aimed at a causal explanation of proficiency using equality of opportunity, or vice versa, but rather at an elaboration of a system-level description of learning outcomes that needs to be explained. The existence or absence of a trade-off between equality of opportunity and the level of learning outcomes is one of the most fundamental questions of education policy because it is directly linked to quite possibly the most fundamental question of the basis of social order in the innate characteristics of the individuals and the effects of the social order on the life paths of individuals.

Unlike previous research, we account for cohort and age differences in learning outcomes and inequality by standardising data by age group and aggregating whole population measures from the standardised age-group-specific scores. Positionality of education and the ability hypothesis are taken into account by introducing a social rank score that measures the positional value of parental education.

The first phase seeks evidence of a trade-off with linear regression of proficiency and equality of opportunity separately for three parental education groups on four different levels of aggregation, aggregating between 5 and 68 birth cohorts. As the opportunity trade-off is a system-level question, related to the causes of between-group differences within systems, the association is analysed across different systems, rather than on an individual level.

The first round of analyses shows a stronger positive association between equality of opportunity and the level of learning outcomes over the long term, especially for the high parental education type, implying an interaction across age groups between equality of opportunity and level of learning outcomes. We analyse the direction of the association across age groups and time or data sets in five different analyses. After demonstrating that equality of opportunity predicts the proficiency of the high parental education type better than vice versa, a model is built that predicts proficiency using equality indicators.

The results show that equality indicators explain a non-negligible part of the achievement differences between countries. The role of anticipatory and contemporary indicators is different for different parental education types. While proficiency of the lower parental education type is strongly associated with contemporary equality indicators, higher parental education type indicators from earlier data significantly improve the prediction. The results imply that the evolution of equality of opportunity may provide advance warning of the direction of the evolution of learning outcomes. Here the results obtained match those from PISA, indicating that, while high proficiency can be achieved with high levels of equality of opportunity, the two usually go together.

After noting that adult data would indicate that developed countries still find themselves on the upward sloping section of the ability-opportunity curve, we turn to seek any traces of education reforms that might be seen in the evolution of equality of opportunity across cohorts.

Our two data sources, IALS and PIAAC, give a largely concordant picture of the evolution of proficiency across cohorts. The pattern also applies to the youngest groups of IALS, implying that experiences in adulthood do not significantly alter the level of equality of opportunity manifest at youth. This makes it feasible to identify discontinuities in adult data that might be associated with education policy reforms. We approach the question by comparing the timing of the observed discontinuities with the timing of education policy reforms. Rough cross-sectional data does not allow us to establish well-defined causality, but if similar education policy reforms are associated with similar cohort discontinuities, it is feasible that the observed cohort effects are treatment effects due to education policy. This would entail that certain types of policy reforms may be associated with cohort differences in realised equality of opportunity.

Some 29 discontinuities are identified that can, by their timing, be associated with notable changes in education policy. While the reforms cannot be analysed in detail, it is noted that some types of reforms seem mainly to be associated with improved equality of learning opportunity, while in some countries similar reforms in the 1990s seem to be associated with declining equality of opportunity.

In presenting the results, we take a brief, slightly more in-depth look at Finland, where the analysis of the evolution of equality of opportunity over time identified a large negative discontinuity in the 1990s, which matches the timing of significant educational reforms. The observation is of interest for several reasons. It would, if the results of our analyses are anything to go by, predict a later decline in learning outcomes, something that Finland has famously observed over the last 13-25 years. In addition, many of the internationally-known characteristics of the Finnish education system have been linked to the reforms of the 1990s and those reforms have been used to explain the rise of Finland to international fame and to the position of a global benchmark. The results suggest that the causes of Finnish success might have been misidentified. The section reviews additional evidence, demonstrating that many other data sources also call into question the story of Finnish rise to prominence due to the reforms of the 1990s. The results suggest that we may have, both nationally and internationally, learnt the wrong lessons from Finland.

The different theoretical and empirical strands of the work are drawn together in ***Section VI***, consisting of ***Chapter 13***, which also presents some paths for future research.

SECTION I - EQUALITY OF OPPORTUNITY

1 Equality of opportunity

1.1 Equality and responsibility

Equality

Dworkin (2002) noted that

> no government is legitimate that does not show equal concern for the fate of all those citizens over whom it claims dominion and from whom it claims allegiance. Equal concern is the sovereign virtue of political community – without it government is only tyranny…

But what does equal concern require? Equality is a universally accepted and universally contested concept, and thus any work dabbling in equality must make it clear what is meant by equality. As Dworkin (1981) states:

> someone who argues that people should be more equal in income claims that a community that achieves equality of income is one that really treats people as equals. Someone who urges that people should instead be equally happy offers a different and competing theory about what society deserves that title. The question is then: which of the many different theories of that sort is the best theory?

The concept of interpretive concept is useful in avoiding naïve literalism when facing such concepts as liberty or equality. Both not only have a non-normative meaning but also a multitude of different normative interpretations. Equality of opportunity is an interpretive concept, as are all the normative concepts that are needed to define it – such as justice, equality, freedom and responsibility. Interpretive concepts do not refer to observable brute facts in the world and are only understandable in terms of each other, as part of a more extensive web of concepts and conceptions. Equality of opportunity, as well as all the key concepts that are instrumental in defining it, has been interpreted in various ways, and the current work attempts to provide a consistent and justifiable interpretation of it.

The distinction between the natural and normative concepts of equality is important. As Cohen (2009) points out:

> If you can reach only the oranges, and I can reach both them and the apples, and I get two oranges and two apples, and you get only two oranges, then you are on the downside of an inequality between us. But if we can each reach both, and I end up with oranges and you with apples, or even if I end up with more of each than you do (because you care less than I do for fruit), then no relevant inequality of distribution holds (unless, what I here assume to be false, the menu was rigged to suit my tastes). It does not matter whether the outcome, in which there is no inequality that reflects unequal opportunity, can or should be called *equality, tout court*, as opposed to *justified* (on egalitarian grounds) *inequality*.

As shall be seen, much has been made of equality in criticisms of egalitarianism that take equality to mean levelling down until all are the same in all respects. While such criticisms are directed against an imaginary foe in the field of normative thought, they gain some semblance of credibility by obscuring the difference between natural and normative interpretations of equality. In the following, inequality will refer to normatively objectionable inequality unless otherwise specified, or unless the

context indicates that we are referring to statistical inequality. Overall, the aim is to be specific enough that the sense in which different interpretive concepts are used is clear to the reader.

Equality of outcomes

The natural place to start the presentation of different conceptions of equality is the equality of outcomes. Two different versions of equality of outcomes may be identified: 1) sameness of outcomes and 2) sameness of the level of outcomes in some common currency. For both, conceptions differ as to what outcomes should be equalised as a matter of justice.

The first position is quickly dealt with, as it is only found in arguments directed against egalitarianism or equality as a value. Murray Rothbard (2000) supplies an example of the type.

> There is one and only one way, then, in which any two people can really be 'equal' in the fullest sense: they must be identical in all of their attributes. This means, of course, that equality of all men—the egalitarian ideal—can only be achieved if all men are precisely uniform, precisely identical with respect to all of their attributes.

No known egalitarian has insisted that individuals should be made identical in every conceivable respect to achieve equality, or argued for absolute equality of everything, with unjust equality being recognised simply from the inequality of a distribution. As Temkin (2017) notes:

> nobody is a radical egalitarian, in the sense of regarding all inequalities as objectionable. After all, there are countless inequalities in the universe — such as the inequality between the number of electrons and neutrons, or between the number of mammals and insects — that are, by themselves, merely descriptive. Accordingly, any reasonable egalitarian will need to seek a deeper explanation than 'mere inequality' to explain why some inequalities are normatively significant while others are not.

The position of equality as absolute sameness seems to arise out of insufficient recognition of the fact that equality is an interpretive concept, and there is a difference between descriptive and normative (unfair) inequality. Indeed, egalitarians that take equality to be a value do not want everyone to be the same any more than libertarians who argue for liberty want everyone to be free to rape and pillage without limit.

Somewhat more interesting are egalitarian positions that take the proper equalisandum to be some aggregate measure that makes all outcomes somehow comparable and measure equality by the distribution of this common currency. They allow a diversity of tastes and preferences to shape the bundles of goods as long as in the bundles are equal in the aggregate. Also, inequality can be measured multidimensionally, thus recognising that individuals have different preferences and ascribe different subjective values to different goods. The generalisation of the principles of equal allocation of goods into a world of bundles of goods leads to the *no-envy principle,* where equality obtains when individuals possessing different bundles of goods do not prefer the bundle of anyone else over their own bundle. The strength of the envy principle lies in taking the valuations of the individuals involved as the basis of valuation of the bundles, instead of relying on a universal value. Perfectly envy-free distribution is of course very demanding, as it is easy to imagine cases where an aspiring musician

talented in physics would envy an aspiring physicist endowed with musical talent. In such cases, the balance of envy can be measured to assess equality of the distribution of goods. (Fleurbaey, 2008)

The literature mostly still aims to find a common metric to measure equality, such as welfare, utility or wealth, even if it is generally recognised that in practice the exercise is one of multidimensional optimisation. When this salient metric can be applied to assess the justice of distribution, without needing any information on how the distribution came about, then the egalitarian theory in question is an *end-state* or *patterned principle of distributive justice*. (Nozick, 1974

Whichever is the principle of equal outcomes, any conception of equality of outcomes needs to specify what outcomes are subject to equalisation. Again, criticisms of egalitarian principles might present the issue as if egalitarians wanted complete uniformity in everything. The other extreme position is represented by entitlement theorists[2], who defend minimal equality of outcomes as formal equality before the law. Extreme positions, defending equality in everything or in nothing, are not to be found in the literature.

Equality in the absence of fair reasons

Equality of at least some outcomes is present in all conceptions of justice because, on a fundamental level, the concept of equality is at the heart of the search for religious, philosophical or political conceptions of justice. Indeed, any normative structure, be it moral or legal, requires consistency, and that consistency is achieved through treating individuals, acts and situations as the same or similar in some cases and different in others. A rule is not a rule if it is different for everyone and in every instance.

The universal requires the non-particular: something that is common to and comparable in all concerned, and that applies to them. It requires a measure: criteria for recognising which entities are to be covered by the comparison, the relevant qualia of the entities that are to be compared and criteria for assessing equivalency. Sameness between humans acts as a basis for treating them equally, and their difference from petunias justifies the differing treatment of humans and flora.

Where there are reasons, there is also the absence of reasons. Generalising from what Arneson (2000) presents in his treatment of moral meritocracy, we may note that the goodness of equality of distribution is entirely a by-product of what matters - namely, that holdings should vary for a reason. Equality is what prevails when there are no reasons for individuals to have different rights, receive different treatment or attain different outcomes. Or, as Temkin (2011) put it, 'in a world where there is no meaningful free will, everyone is *equally* deserving—as well as equally *un*deserving!—since no one morally *deserves* anything at all.' As such, equality does not require any egalitarian principle to demand equalisation; it is what is left when there are no reasons for things to be any particular way. From this perspective, it does not become a default because it is the primary principle to be applied, but simply because an absence of reasons for differences implies equality.

This is also why the question 'why equality?' is equivalent to 'equality of what?' (Sen, 1992) All conceptions of justice have things that people should be made equal in respect to because there

[2] Broadly, entitlement theory can be taken here as a synonym of right libertarianism, covering such thinkers as Nozick, von Hayek, von Mises and Rothbard. While only Nozick refers to the position as entitlement theory, all share a very similar understanding of the principles of justice when it comes to equality. In addition, the use of the term entitlement theory is a response to the apt criticism of Cohen that too much is granted to libertarians when their position is presented as a defense of liberty. (Cohen, 1995)

are no justifiable reasons for differences between them. Nozick's (1974) critique of the Rawlsian equality of primary goods argues for equality of libertarian rights. Sen (1992) has emphasised that utilitarianism, even when it is indifferent to the separateness of persons, uninterested in the distribution of utility, and only focused on the maximisation of aggregate utility, still 'guarantees that everyone's utility gains get the same weight in the maximizing exercise.' All conceptions of justice share the approach that there are reasons why things are one way and not another, and that some reasons are fair, and some reasons are not fair. This is how comparative fairness and equality dovetail or can in some cases be considered two sides of the same coin. (Temkin, 2017)

Thus, it is conceived that all theories of justice answer Sen's question: 'equality of what?' Theories of justice are also inherently comparative and, correspondingly, any conceivable theories of justice insist on the comparative fairness of equality of individuals in the relevant *equalisandum*. Indeed, 'equality represents a distinct concern of significant moral value, a concern about comparative fairness'. (Temkin, 2017) It is shared by all theories of justice because they are theories of what is fair and of the interpersonal or relational requirements of fairness. The disagreement is over the terms on and space in which equality should be assessed. The history of the discussion on equality is the history of the discussion on the proper *equalisandum*.

A multitude of different *equalisanda* have been proposed and defended, and a thorough presentation is unnecessary for the purposes of this work. We may note the variety of responses to what individuals should be made equal in and the importance of the answer to this question by quoting Temkin (2011):

> host of candidates have been championed, including: income, resources, primary goods, wealth, power, welfare, opportunity, needs satisfaction, capabilities, functionings, rights, and liberties. It is difficult to exaggerate this topic's importance, since equality of one kind often *requires* inequality of another. For example, equality of income may correlate with *inequality* of need satisfaction between the handicapped and the healthy, and vice versa.

Responsibility and freedom

But there is also the other side of the coin. Reasons for treatment or outcomes to be different are present in all theories of justice. We may begin with the link between individual freedom and responsibility. Responsibility arises as a necessary consequence of individual freedom. The link is almost universally accepted in both retributive and distributive justice, and strongly emphasised by writers such as Friedrich von Hayek (2013), who noted:

> Liberty not only means that the individual has both the opportunity and the burden of choice; it also means that he must bear the consequences of his actions and will receive praise or blame for them. Liberty and responsibility are inseparable.

Here the justification of normative responsibility follows from the possibility of causal responsibility. This link ties responsibility and freedom together. Freedom requires the possibility to choose or act in different ways. Actions have consequences and individuals can be held responsible for those consequences.

The responsibility of the individual follows from making a choice between counterfactuals. When there is no alternative, no choice to be made, the question of moral judgement does not arise.

The agent must have the capacity, in the specific case, to change the normatively assessed outcome. It is not enough to have the capacity to do so in general, or in some other cases if, in the case at hand, the agent has faced a hard obstacle that has prevented them from changing the outcome.

Later on, we will note that individuals are sometimes held responsible for causes beyond their control.[3] For now, it is enough to note that founding moral responsibility on causal responsibility is what Temkin (2011) refers to in referring to the robust conception of responsibility.

> People must, in a robust sense, be responsible for their characters for them to be morally deserving of being well or poorly off in virtue of their characters, and hence for it to be a matter of absolute justice that they be well off in direct proportion to the degree to which they are virtuous.

Merit and desert

Responsibility for the exercise of one's freedom is closely linked to the concept of merit or desert. These concepts are sometimes used interchangeably, sometimes a distinction between the two is made. The distinction is linked to the two perspectives of merit that are sometimes left unspecified in the discussion, but can be seen in the Cambridge English Dictionary (2019) definition of merit as 'the quality of being good and deserving praise', which reflects the old French origins of the concept in the word *merite*, referring to both excellence and moral worth. The crucial issue is, obviously, that excellence may not always deserve praise and sometimes praiseworthiness is achieved without excellence.

Here we are concerned with the concept in the second sense, as moral worth deserving praise. In this second sense, merit is often used interchangeably with desert, which emphasises the quality of being deserving. More precise distinctions are also possible, whereby desert is reserved as a label for the quality of deserving, whilst merit refers to the quality of excellence against some set of criteria but does not (necessarily) carry the connotation of providing desert. Here it is not necessary to go into further nuances of terminology, as the crucial issue is that merit or desert is intimately associated with the concept of responsibility. An individual may and cannot acquire merit by things she does not influence, any more than a king may be held responsible for the rise of the sun he commands to rise. (Saint-Exupéry, 1943) Merit is the normative assessment of individual action, and individuals might merit being rewarded for their actions by increased welfare or a jail sentence. The crucial issue is that merit only follows from choices.

> Egalitarians generally regard differently those differences in individual capacities which are inborn and those which are due to the influences of environment, or those which are the result of 'nature' and those which are the result of 'nurture.' Neither, be it said at once, has anything to do with moral merit. Though either may greatly affect the value which an individual has for his fellows, no more credit belongs to him for having been born with desirable qualities than for having grown up under favourable circumstances. (Hayek, 2013)

The merit that is acquired by choice may not be collective or inborn. As individual merit can only be attached to something that an individual can be causally responsible for, normative responsibility

[3] See Section II – Innate ability and responsibility

becomes fundamentally individual. Hayek (2013) correctly dismisses collective responsibility in cases where individuals have not acquired individual responsibility by taking part in collective action.

> Freedom demands that [an individual] […] be responsible for only his own actions […] not for those of others who are equally free. […] In a free society there cannot be any collective responsibility of members of a group as such, unless they have, by concerted action, all made themselves individually and severally responsible.

This combination of freedom and responsibility for the consequences of the exercise of that freedom has been seen as the basis of a free society. Hayek (2013) maintains that such a society will not survive if individuals do not accept the justice of bearing the consequences of their actions.

> A free society will not function or maintain itself unless its members regard it as right that each individual occupy the position that results from his action and accept it as due to his own action.

The area of responsibility is thus the area where 'the individual has both the opportunity and the burden of choice', but also bears 'the consequences of his actions' (Hayek, 2013), where individual differences in outcomes are to be carried by individuals as they have arisen out of 'fault or choice of their own' (Temkin, 1993) and where an individual *can* be held responsible for outcomes that *do* 'appropriately reflect choices that he has made or is making or would make.' (Cohen, 2011)

Currency of responsibility

Within egalitarianism, there is a variety of positions on what individuals should be held responsible for. The main perspectives on the object of responsibility have been control and choice or preferences. In the control perspective, individuals should be held responsible for the outcomes under their control, in the choice/preference perspective for the consequences of actions that they have chosen. In philosophically oriented discussion, the control view has been defended by e.g. Arneson, Cohen and Roemer, while the choice view has been defended by Rawls, Dworkin and Fleurbaey.

Most empirical approaches have explicitly or implicitly adopted the control view of responsibility, holding individuals responsible for the causes of outcomes they effectively control while Fleurbaey and Maniquet have defended the preference cut. (Ferreira & Gignoux, 2013; Roemer, 1998; Roemer, 2012; Roemer & Trannoy, 2016, Fleurbaey,2008) This prevalence of the control view in the field of economics might be understood as an immanent criticism of positions that defend narrowly construed procedural justice while justifying responsibility as a necessary prerequisite and consequence of freedom within economics, held by economists such as Hayek and von Mises. Complementarily, the control view might be understood as a conception of equality that corresponds to the methodological individualism widely adopted in economics in general, and to economics as focused on incentives faced by individuals in conditions of scarcity. The link between causal and moral responsibility is natural because the sphere beyond causal influence is also indifferent to individual incentives.

Assuming free will brings the control and choice views very close to each other, as Fleurbaey contends, but also leaves the fundamental difference in that the choice view holds individuals responsible for their unchosen preferences, which they identify with, while the control view does not. Fleurbaey argues that the control view 'leads theories of justice into a metaphysical dead end' and

presents a valuable criticism of the control view by focusing on the practical difficulty of identifying choice, the difficulty of fitting genuine choice into models of social science that explore decision making as optimisation exercises and, most importantly, by questioning the principle because of its dependence on the assumption of free will. (Fleurbaey, 2008)

Dworkin would hold individuals responsible for the preferences that they have and identify with, while Cohen would place the cut at where a choice has been made so that individuals are not held responsible for preferences that they might have due to the circumstances under which they were born or grew up. In Dworkin's conception, the lifting is done by the requirement that the individual identifies with her preferences, while Cohen (2011) notices that an individual may well identify with and not regret a preference, but regret the fact that the preference is cause for hardship or discrimination under current social arrangements.

While Fleurbaey's (2008) criticism is in many ways valid, the assumption of free will is defensible under any conception of justice, as all conceptions of justice require a counterfactual. Technically, Fleurbaey's solution is excellent, because preferences and choices are always there, even if they are determined by hard determinism and free will does not exist. But it sacrifices the fundamental normative underpinnings of the approach. It seems difficult – well-nigh impossible – to imagine a morality that would be able to function without free will. There is no right or wrong if no one can ever choose anything. Temkin's (2011) observation above is relevant here as well. '[I]n a world where there is no meaningful free will […] no one morally *deserves* anything at all.' For now, we do not understand the question of free will very well – general experience would lead us to believe that it exists, but we do not understand how it arises from mechanistically operating nature.

Because throwing free will overboard would seem to sacrifice the whole project of justice, or at least make it irrelevant, it is hard to see what would be saved by adopting the choice perspective to avoid the question of free will, besides technical identifiability of responsibility characteristics.

The conflict between equality and freedom

This link between freedom and responsibility is behind the often-presented view that freedom and equality conflict or form a trade-off. The rationale is simple: liberty requires responsibility and individuals are not held responsible if everyone attains the same outcomes irrespective of their actions. If outcomes are redistributed between individuals, some individuals will lose outcomes that they have contributed to bringing about. Freedom is limited, as the range of outcomes of free individual action is limited.

This is the core of Nozick's (1974) most general, influential and forceful critique of the Rawlsian conception of justice as fairness. (Rawls, 2009) Nozick makes a distinction between *current time-slice principles* or *end-result principles*, which 'hold that the justice of a distribution is determined by how things are distributed (who has what) as judged by some structural principle(s) of just distribution'. He contrasts them with *historical principles*, which consider that the justice of a distribution 'depends upon how it came about.' Rawlsian principles of justice are the primary target of his criticism of such end-result principles.

Nozick (1974) reasons that if patterns of outcomes determine justice, then

> to maintain a pattern one must either continually interfere to stop people from transferring resources as they wish to, or continually (or periodically) interfere to take from some persons resources that others for some reason chose to transfer to them.

Thus, 'patterned principles of distributive justice necessitate redistributive activities', and as 'seizing the results of someone's labour is equivalent to seizing hours from him and directing him to carry on various activities', this makes 'them a part-owner of you; it gives them a property right in you.' Thus, the end-state principles abridge individual freedom, emptying it of content. Nozick (1974) argues that

> distributional principles do not give people what entitlement principles do, only better distributed. For they do not give the right to choose what to do with what one has; they do not give the right to choose to pursue an end involving (intrinsically, or as a means) the enhancement of another's position.

Freedoms

Here it is useful to distinguish between different types of freedom. Isaiah Berlin (1958) made famous, following Constant (2010), the distinction between negative and positive liberty. Negative liberty is conceived as freedom from interference and obstruction by others.

> If I am prevented by others from doing what I could otherwise do, I am to that degree unfree; and if this area is contracted by other men beyond a certain minimum, I can be described as being coerced, or, it may be, enslaved.

Positive liberty, in contrast, is conceived as the freedom to something.

> The 'positive' sense of the word 'liberty' derives from the wish on the part of the individual to be his own master' I wish my life and decisions to depend on myself, not on external forces of whatever kind. (Berlin, 1958)

Another critical distinction, one that is much less well recognised, is between what Cohen refers to as *neutral* and *rights accounts of liberty*. In the *neutral account*, any interference with an individual diminishes freedom. (Cohen 1995) But it then follows that, if liberty is understood wholly in negative terms, 'egalitarian policies can only redistribute liberty: they cannot be shown to diminish it.' (Steiner, 2016) The reason is simple: individual action is curtailed when the rights of other individuals are defended. If any interference with the actions of an individual is interpreted as a limitation of freedom from interference, then any interference to safeguard the rights of one trespasses on the liberty of another and thus simply reallocates negative liberty between individuals rather than affecting the aggregate amount of liberty involved.

The need to allow for rightful interference with the freedom of actions of others has led to what Cohen refers to as the *rights account of freedom*. Nozick (1974) states that

> other people's actions place limits on one's available opportunities. Whether this makes one's resulting action non-voluntary depends upon whether these others had the right to act as they did.

In this Nozick echoes Hayek (2013), who wrote that

> the conception of freedom under the law […] rests on the contention that when we obey laws, in the sense of general abstract rules laid down irrespective of their application to us, we are not subject to another man's will and are therefore free.

This strategy successfully removes the stigma of limiting freedom from the action that removes a trespasser from a field by postulating that, as the owner had the right to exclude the trespasser from her property, this removal did not limit the trespassers' freedom. Cohen (1995) demonstrated that this success comes, for entitlement theorists, at a steep price:

> Suppose that we overrule ordinary usage and we say, with Nozick, that rightful interference with someone's action does not restrict his freedom. It cannot then be argued, without further ado, that interference with private property is wrong because it restricts freedom. For one can no longer take for granted, what is evident on a rights-neutral ordinary language conception of freedom, that interference with private property does reduce freedom. On a rights account of what freedom is one must abstain from that assertion until one has shown that people have moral rights to their private property. Yet libertarians tend both to use a rights definition of freedom and to take it for granted that interference with his private property diminishes its owner's freedom. But they can take the latter for granted only on the rights neutral account of freedom, on which, however, it is equally obvious that the protection of private property diminishes the freedom of non-owners, to avoid which consequence they adopt a rights definition of the concept. And so they go, back and forth, between inconsistent definitions of freedom, not because they cannot make up their minds which one they like better, but under the propulsion of their desire to occupy what is in fact an untenable position. Libertarians want to say that interferences with people's use of their private property are unacceptable because they are, quite obviously, abridgements of freedom, and that the reason why protection of private property does not similarly abridge the freedom of non-owners is that owners have a right to exclude others from their property and non-owners consequently have no right to use it. But they can say both things only if they define freedom in two incompatible ways.

Thus, it is essential to recognise the distinction between the neutral and rights accounts of freedom. The concept of equality of opportunity requires thoroughly intertwined conceptions of equality and freedom.

1.2 Combining equality and responsibility

Tale of two principles

Equality of opportunity holds that individuals are entitled to equal opportunities as a matter of justice but are responsible for how they use the opportunities available to them. Temkin (1993) has condensed this combination of equality and responsibility in his 'canonical formulation of egalitarianism' as:

> it is morally bad – unjust and unfair – if some people are worse off than others through no fault of their own.

He specifies that by fault he means 'fault or choice', so that individuals may be held responsible for the outcomes that are due to their own choices and actions, but not for those that are not. This formulation contains the principles of horizontal and vertical equity, where horizontal equity is concerned with equality of equals and vertical equality with simply inequality of unequals. Horizontal equality requires that individuals that are equal in normatively pertinent respects receive equal treatment or outcomes. Vertical equality, on the other hand, requires that relevant differences in actions or characteristics lead to correspondingly different outcomes or treatment. Temkinian canonical egalitarianism thus takes it as unfair if individuals are unequal through no fault or choice of their own. By implication, it considers it unfair if individuals are equal even if they merit being unequal in some respect.[4]

These two aspects of equity constitute the principles of equality and responsibility, respectively. These principles of distributive justice correspond to the principles of retributive justice that wrongful acts merit punishment and that it is wrong to punish the innocent. Both principles are related to the two famous Rawlsian principles of justice as fairness:

> First: each person is to have an equal right to the most extensive basic liberty compatible with a similar liberty for others. Second: social and economic inequalities are to be arranged so that they are both (a) reasonably expected to be to everyone's advantage, and (b) attached to positions and offices open to all. (Rawls, 2009)

These principles are usually known as the liberty principle and the difference principle. For our purposes, we emphasise the aspect that the first pertains to giving something to each individual in equal measure while the latter pertains to what is needed to justify individuals possessing something in unequal measure. Rawls specifies liberty as the currency to be equalised, but the principle of distribution is the same as with Temkin, that it would be unjust for anyone to have less or more than another. As for the correspondence of the Rawlsian difference principle with the responsibility principle as defined above, canonical Temkinianism emphasises that it would be unjust for individuals to be equally well off if they merit otherwise, while Rawls focuses on the conditions under which individuals may merit otherwise. Thus, the Rawlsian principles of justice as fairness correspond to Temkin's formulation of the principles of canonical egalitarianism.

Different clothes of the two principles

Both forms of the principles translate into determining what individuals are held and not held responsible for. When individuals are held responsible, they may acquire the merit, positive or negative, that justice requires, according to vertical equality, to be rewarded by corresponding differences in outcomes. When individuals are not held responsible, there are no reasons that would justify putting individuals on different vertical planes, and the principle of horizontal equality prevails.

At the core of equality of opportunity, including its empirical operationalisations (Roemer, 1998; Roemer & Trannoy, 2015), is the division of determinants of outcomes into *circumstances* and

[4] Of course, it may be noted that, while the principle takes it as morally forbidden for the equally deserving to be unequal, it is not clear that the unequally deserving would be required to be unequal, in contrast to merely being permitted to be unequal.

effort. In this distinction, circumstances are the realm beyond individual control, imposed on the individual and thus beyond individual responsibility, whilst effort refers to individual choice and action that the individual can be held responsible for and acquire merit by. This terminology draws attention to the question of responsibility, linking just compensation and rewards to the ability of the individual to choose. Cohen argues that equality of opportunity should remove

> disadvantage for which the sufferer cannot be held responsible, since it does not appropriately reflect choices that he has made or is making or would make. (Cohen, 2011)

While the terminology of effort-circumstances is common in the literature, here the concept of *responsibility characteristics* is used to refer to legitimate grounds for differences in individual outcomes. In the hope to align technical usage with everyday usage, and the language of statistical analysis with that of philosophy, *effort* is here reserved for individual responsibility characteristics that individuals can directly exercise control over. The *reward principle* refers to the differences in outcomes justified by the responsibility characteristics.

A third perspective on the same divide comes through the outcomes that are distributed. Distribution of outcomes acts through the *reward principle* and the *compensation principle*. The *reward principle* is concerned with merited rewards for differences. The rewards may be positive or negative, depending on the nature of the differences to be rewarded. Different reward principles may reward individual achievement to different degrees, even when they all share the approach that distribution somehow follows merit. The reward principle sits in contrast to the *compensation principle,* which is concerned with distribution in the 'before' phase when all that is to be distributed is circumstances. (Fleurbaey, 2008) While the reward principle ensures that responsibility is associated with just rewards, the compensation principle refers to compensatory actions to be taken when individuals face unmerited 'rewards' not associated with responsibility characteristics.

The metaphor of fair rules for a fair race may be emphasised by talking of 'before' and 'after', as in before and after the race begins. Substantive equal opportunity aims at a level playing field in the 'before' state, which covers circumstances beyond individual choice or control just as well as all principles of justice applicable to all in equal measure. 'After' refers to outcomes after individual choices and effort, and individuals may be held responsible for the consequences of their choices and actions. Equality should obtain in the 'before' phase, but differences in the 'after' phase are not necessarily unjust. (Roemer, 1993, 1998, 2006; Sen, 1992; Temkin, 2013) The difference between 'before' and 'after' is the sphere of individual choice, effort and responsibility.

> There is, in the notion of equality of opportunity, a 'before' and an 'after': before the competition starts, opportunities must be equalized, by social intervention if need be, but after it begins, individuals are on their own. (Roemer, 1998)

All of these perspectives distinguish between the domains of responsibility and of circumstances. In the domain of individual freedom, responsibility for the consequences of the use of that freedom prevails. In contrast, individuals are not held responsible for their circumstances. All three are entirely equivalent so that *compensation* principles apply to *circumstances* in the *before* state, which is turned into after by individual *effort*, rewarded by *reward* principles. The universality of the distinction is essential, as it means that, when the compensation principle is not applied, all the rest is the application of the reward principle and vice versa. Reward principles exist in many forms, but Fleurbaey

(2008) distinguishes between *liberal* and *utilitarian* rewards, the former of which represents the results of responsible action/choice under laissez-faire policy, while the latter refers to reward principles that would maximise aggregate outcomes.

Moral meritocracy

It is noteworthy that the basic egalitarian concern that 'it is bad – unjust and unfair – for some to be worse off than others through no fault or choice of their own' (Temkin, 1993) is entirely consistent with distribution of outcomes fully on the basis of merit as it

> is fully compatible with (2): It is morally bad –unjust and unfair – if some people are as well off as others through no merit of their own. Taken together, (1) and (2) are compatible either with a qualified affirmation of equality or (modulo a qualification to be noted) with a principle of moral meritocracy (3): It is morally good – just and fair – that each person be exactly as well off, by comparison with others, as she deserves to be. The more deserving one is, the better off one should be. (3) would have it that people who are equally deserving should be equally well off, but the goodness of equality of distribution here is entirely a by-product of what really matters, namely, that well-being should vary with desert. (Arneson, 2000)

Thus, Temkin's canonical formulation of egalitarianism does not 'unequivocally affirm any egalitarian principle'. Instead, it is consistent with *moral meritocracy*, which distributes outcomes solely based on moral merit. Any egalitarianism involved only arises because equality prevails in the absence of reasons for an unequal distribution. If non-merit causes should not exercise any influence on the distribution of outcomes, the distribution of merit fully determines the outcome distribution. In a moral meritocracy, the only just reason for differential outcomes is merit, and we only get equal distribution when merit is silent.

> It is not that according to moral meritocracy there is some initial reason to equalize the distribution of good fortune, a reason that might be outweighed by subsequent events. There is no initial reason to equalize at all. It is as though just before the start of a horse race one observed that there is some reason to award every horse the prize for finishing first, since none has so far run faster or slower. But none has run at all, so the principle of distribution according to merit is silent. (Arneson, 2000)

The two dimensions of merit come into play in the definition of moral meritocracy. Temkin's (2013) definition of the *equal opportunity merit principle* is effectively equivalent with the Rawlsian *fair equality of opportunity* in that it judges merit not by moral merit, but by merit as goodness of fit to the demands of the position in question. Thus, it is open to the same criticism that Arneson levelled against the Rawlsian conception. This matter will be further discussed in chapter 5 on Innate ability.

Equality of opportunity is not everything

Some take moral meritocracy to be a non-egalitarian principle or one that is not sufficiently egalitarian. These criticisms take the stance that is misleading to call it equality as it simply safeguards individuals the right to become unequal. Competition can be fair in terms of *procedural* and *background fairness* yet result in however unequal a distribution of outcomes. Indeed, equality of opportunity is

consistent with any distribution of outcomes as long as equality obtains in the distribution of opportunities. This principle is captured by Roemer:

> what society owes its members, under an equal-opportunity policy, is equal access; but the individual is responsible for turning that access into actual advantage by the application of effort. (Roemer, 1998)

Temkin captured the same crucial feature of equality of opportunity:

> if one could ensure that people's starting points fully met the egalitarian's criteria for equal opportunity – so, people were dealt similar hands in terms of talents, temperament, individual responsibility, and life prospects – and if, in addition, one could rectify the influence of luck on people's lives – so, ultimately, each person was responsible for how well they fared – then, one could be satisfied with the outcome in terms of comparative fairness regardless of whether it involved ex-post equality. (Temkin, 2013)

Thus, when full equality of opportunity obtains, all individuals face the same expected value of outcomes irrespective of their circumstances, as the distribution of outcomes follows from the distribution of responsibility characteristics. The approach does not determine how unequal the distribution of outcomes should be. Two distributions of outcomes are purely equivalent for the realisation of equality of opportunity if there is no difference between the distributions in the association between responsibility characteristics and outcomes. In this sense, equality of opportunity is interested in ordinal rankings and assesses the ordinal ranking of responsibility characteristics against an ordinal ranking of outcomes. Equality of opportunity fails to obtain if the two ordinal rankings do not match.

But, as there is no commensurate way to measure responsibility and outcomes, the distribution of responsibility characteristics implies nothing in terms of the cardinal distribution of outcomes (Fleurbaey, 2008). This raises into consideration principles of justice that limit just distributions even when the distribution is entirely a function of individual choice and effort. As Roemer has noted: 'considerations outside the realm of equality of opportunity must be brought to bear to decide upon how much inequality with respect to differential effort is allowable.' (Roemer, 1998)

Jacobs refers to these considerations as *stakes fairness* and would include them within the concept of equality of opportunity. He sees that the discussion on equalities of opportunity has moved from one-dimensional (procedural fairness) to two-dimensional (procedural & background fairness) conceptions of equality of opportunity and advocates a move to a three-dimensional conception that includes stakes fairness. (Jacobs, 2003)

The question of stakes fairness may be divided into two considerations: 1) is it fair to hold something, like access to justice or health care, as a stake and 2) is the stake in proportion to the difference in merit? Jacobs emphasises the former dimension – that it would be unjust to allocate some things competitively, even if the competition were fair in terms of procedural and background fairness. We would baulk at the idea of fair competition for treatment before the law, with winners of an entirely fair meritocratic competition receiving a better position than the losers. (Jacobs, 2003)

The other aspect, the proportionality of reward to merit justifying the reward, was referred to by Roemer above. He has been worried that (two-dimensional) equality of opportunity may be insufficiently egalitarian and draws attention to the fact that we would need, and do not have, a theory of just rewards to effort. (Roemer, 2012) Such a theory of just rewards to effort is what Temkin refers to as the difference between *absolute* and *relative proportional justice*. Absolute proportional justice

would give an absolute scale on which effort/moral merit should be rewarded by outcomes. Relative proportional justice, in turn, would simply demand that the more virtuous fare better than the less virtuous, but would not necessarily demand that sinners must fare badly, or saints must fare well. (Temkin, 2011)

Such questions are of interest independent of the realisation of equality of opportunity. Even in circumstances where full two-dimensional equality of opportunity is realised, it may be worthwhile to assess how differences in outcomes match differences in effort. In the area of material wellbeing, the question is what extent of income or wealth differences may be justified by differential effort. In education, we may imagine a system where full equality of opportunity is realised and effort can be correctly identified, and effort is used to identify persons who receive more education. The question then is, how steep may the gradient of investment in education be as a function of effort? Can, for example, the low-talent or low-effort individuals be left practically without education so that they never learn to read, write or count, even if they could do so?

While Jacobs (2003) makes a significant observation when he recognises the absence of stake fairness as a shortcoming of two-dimensional equality of opportunity as a theory of justice, it is less clear that it would be helpful to incorporate the dimension of stakes fairness into the concept of equality of opportunity – especially as Jacobs conceives of stakes fairness very broadly, as one should. The principles by which some things should be left out of the goods to be distributed according to equality of opportunity bring into play principles of prioritarianism and sufficientarianism, taking thus different aspects of justice under the umbrella of equality of opportunity. By including the question of the relationship of effort and reward, it also brings in a theory of just rewards to merit. That approach turns equality of opportunity into a full-fledged theory of justice, a direction that has also been opposed in the empirically-oriented theoretical literature. (Roemer & Trannoy, 2016)

Thus, rather than turning it into the theory of distributive justice, it is here retained as one aspect of distributive justice, and the concept is used to refer to a two-dimensional, substantive conception of equality of opportunity that covers procedural and background fairness.

Stakes fairness and principles of reward and compensation

The question of stakes fairness can also be made using the concept of reward principles. It has been demonstrated that while the compensation (eliminating inequalities due to circumstances) and the reward principles (rewarding effort) are in theory perfectly compatible, they tend to conflict when they are given more precise form. The compensation principle can be combined with different reward principles that produce different results in terms of dividing outcomes as due to circumstances and effort.

Views have also been presented that the reward principle should be asymmetrical, holding individuals more responsible for their choices that turn out well than for choices that turn out badly. (de-Shalit & Wolff, 2011) This approach is echoed by Temkin (2011), who reminds us, making reference to Kant, that ‘one must remember that full responsibility for one’s choices doesn’t entail full responsibility for one’s predicament.’ This asymmetry may take several forms. The possible losses for responsibility may be capped (either relative to the baseline or because the end result is too bad), their scale may be affected by, e.g., taxes so that real gains are larger relative to nominal ones than losses, or their scope is to be limited so that individuals are responsible for fewer sources of misfortune than fortune. (de-Shalit & Wolff, 2011)

For our purposes, it is sufficient to limit ourselves to the two main reward principles that Fleurbaey has named the liberal and utilitarian reward principles. The *liberal reward principle* refers to the principle by which 'no further redistribution should be performed beyond what is required by the compensation principle'. The *utilitarian reward principle* is based on the idea that there is no aversion to inequality for the individual differences that individuals are responsible for, and thus the welfare function is identical to the so-called utilitarian welfare function, which maximises aggregate welfare without consideration for the distribution of welfare. The liberal reward principle may be considered a half-way house between egalitarianism and libertarianism, and the utilitarian reward principle a half-way house between egalitarianism and utilitarianism. The first implies laissez-faire in a hypothetical situation where 'all individuals enjoy equal circumstances and are fully responsible for their differences' while the latter implies that redistribution may surpass the level demanded by the compensation principle if such redistribution improves aggregate welfare. (Fleurbaey, 2008)

As this study is interested in identifying the level of equality of learning opportunity, and learning outcomes are inalienable, it is natural here to assume the *liberal reward* principle, with laissez-faire when circumstances have been equalised. The naturality of the liberal reward principle follows from inalienability because inalienability makes learning outcomes impossible to redistribute and thus the reward function – the relationship between effort and outcomes – is beyond our control. For outcomes that are fundamentally social, like the distribution of material well-being, we may agree on a reward principle of how effort is to be rewarded, and some societies might choose steeper correspondence slopes between effort and outcomes than others, while still applying moral meritocracy. When the reward function is natural, it is beyond intervention, and thus we may not arbitrarily determine the correspondence between effort and outcomes. Thus, the object of our interest means that while some conceptions of justice, such as luck egalitarianism, would require that individuals be compensated for their lower talents (Lippert-Rasmussen, 2011), they would not be compensated in kind, but in other, alienable and redistributable resources. But it is essential to note that while we are here limited to equality of opportunity, even if full (bi-dimensional) equality of opportunity were observed to obtain, it is not clear that justice would prevail.

Equity and equality

The fact that equality of opportunity does not, even for the supporters of the conception, represent a full set of equality-related requirements that society should fulfil, allows us to clarify why the term equity is not present on these pages. The first key feature of equity as a concept is that it passes normative judgement. The Oxford English Dictionary (2019) defines equity as the quality of being fair and impartial. In the realm of education, the Handbook of Measuring Equity in Education takes 'equity to mean that a distribution is fair or justified.' (Cameron, Daga, & Outhred, 2018) As such, equity carries the idea of fairness and normative justification.

Now, there are two approaches as to how equity, as a concept, relates to the norms that are used to pass judgement on anything to deem it equitable. One strategy is to consider that equity, as such, carries no information on the norms in question. As Cameron, Daga, and Outhred (2018) state: 'equity involves a normative judgement of a distribution, but how people make that judgement will vary.' Following Jacob and Holsinger (2008), they define that equity in education 'considers the social justice ramifications of education in relation to the fairness, justness and impartiality of its distribution at all levels or educational sub-sectors.' This type of *non-committal equity* uses equity as a technical

term that refers to the relationship of facts to any set of norms that define equity. It is a vessel for content, an interpretive concept without an interpretation. In non-committal usage, equity is synonymous with justice, fairness or impartiality, not a criterion of any of them. It is an empty vessel, to be filled by criteria but not making a commitment to any set of criteria.

The other approach is to consider that equity contains the norms that define something as more or less equitable. This type of *committed equity* represents an interpretation of the substantial content of equity as a norm. In that usage, equity refers not to anything considered fair by any normative system, but things that are considered fair and impartial according to one defined normative system that has been adopted. In this usage, equity includes the relevant criteria of, e.g. equality of opportunity (Temkin, 2016) or an acceptable minimum level of outcomes (Frankfurt, 2015). As Barry (1990) has noted, 'if the principle is taken to include within itself criteria for determining what makes people "equal" and what makes them "unequal", then it swallows all other comparative distributive principles.'

Thus equity, as a term referring to the match between norms and state of affairs, is either empty of normative content or very heavy with it. In the former case, the term is of no particular analytic use. In the latter, it acts as a repository for any normative commitments that have been made and need to be made explicit in any case, as the concept of equity does not contain any widely agreed-upon normative content besides fairness of the outcome. In practice, the term equity is made to do some of the heavy lifting that was done above by dividing inequalities into unfair and fair inequalities. For example, Levin (2003) notes that

> There is general agreement that the aim of public policy cannot and should not be equality in the sense that everyone is the same or achieves the same outcomes – a state that appears to be both impossible and undesirable. Rather, a commitment to equity suggests that differences in outcomes should not be attributable to differences in areas such as wealth, income, power or possessions. The question is then of what state or degree of inequality is acceptable.

Levin uses the concept of equality as an alternative to a narrowly conceived equality of outcomes. Equity as a term is entirely unnecessary to make this point, as the Levinian intuition considers the common conception of equality of opportunity. What he terms equity is a combination of equality of opportunity and of Frankfurt's standard of sufficiency, i.e. setting an acceptable minimum level of outcomes.

This is, of course, apparent in classical definitions of equity in the literature as well. When it has been defined that 'the principle of equity is that equals should be treated equally and unequals unequally' (Barry, 1990), a commitment has been made to a definition that makes a distinction between qualia relative to which individuals are to be considered or treated as equals and those who merit unequal treatment. This distinction matches the distinction between circumstances and responsibility characteristics and is very close to the conception of equal concern as present in Dworkin (2000) and in Cohen's (2011) basic egalitarian concern.

For the following analyses, the term equity is unnecessary and potentially harmful, as it does not bring any new analytic or conceptual apparatus into play but risks confusing concepts.

2 Opportunity

When we agree that we want to reward individuals for their efforts, but not for their circumstances, we end up with equality of opportunity as the expression of canonical egalitarianism. The trouble is that opportunities are 'inherently unobservable' because 'they are, by definition, a set of hypothetical options'. (Ferreira & Peragine, 2015) The roads diverge in the woods, and when choosing our path, we do not know where it may lead. After choosing - the roads not taken are there no more.

2.1 Opportunity, counterfactuals and causal models

Two steps to what might have been

How would we observe opportunities or assess whether they are equal or not? Two main avenues are open. Approaches in procedural justice focus on the formal criteria of the paths diverging in the woods. Some of the paths are guarded by uniformed men and inequality of opportunity presents itself when, on orders from above, they prevent some wanderers from embarking on some of those paths. If we wish to consider all forms of discrimination, even the forms that we are not fully aware of, this is not sufficient. *Substantive equality of opportunity* requires not only *procedural justice* but also *background justice*, that the competition takes place on a level playing field. (Jacobs, 2003) We may try to measure the levelness of the playing field directly, but there we always suffer from unobservables if and when we are unaware of how the playing field affects the outcomes of the game. To take this approach, we need a model of how the world works to assess opportunities as counterfactuals, as what might have been.

Thus, a strategy of taking procedural justice as formal equality of opportunity is built on a Procrustean definition of opportunity. It limits the concept of opportunity to a defined set of causes that contribute to the distribution of outcomes. Any cause of outcomes that is not on the list – however much beyond individual control and contributory to the eventual outcomes – is not considered to be an opportunity. This sets up a division between formal opportunity, the things on the list, and actual opportunity, including all the other circumstances bearing on the outcomes. This amounts to making significant normative commitments without justifying them normatively. This is done by accepting – through a narrow definition of opportunity – some causes beyond individual control or choice as just causes of outcomes. As this amounts to bypassing the crucial questions of normative justification, we follow the literature in adopting the indirect approach to identifying opportunity.

The indirect approach to identifying opportunity makes it necessary to build a causal model of the world to produce the counterfactuals that are assessed. This causal model is inferred by observing the variation of outcomes with different observable circumstances. The current identification strategies build on the understanding that the causes of outcomes are divisible into circumstances and effort. This makes inequality of outcomes a combination of the distributions or inequalities of opportunities and effort. (Marrero & Rodríguez, 2013) Empirical strategies seek to disentangle these two distributions from each other by observation of 'joint distributions of circumstances, effort and outcomes'. (Ferreira & Peragine, 2015; Temkin, 2013).

The combination of inference and measurement gives rise to the two-stage nature of measuring inequality of opportunity, which distinguishes it from the measurement of inequality of outcomes. It is often referred to as a two-step procedure:

> first, the actual distribution is transformed into a counterfactual distribution that reflects only and fully the unfair inequality, while all the fair inequality is removed. In the second step, a measure of inequality is applied to [counterfactual distribution][5] (Ferreira and Peragine, 2015).

These two steps can be called, following Roemer and Trannoy (2016), the *estimation* and *measurement* stages.[6] The estimation phase is devoted to the estimation of the mechanism producing the outcomes, to identify unjust causes, while the opportunity distribution is measured in the second phase. First, we will present how a causal model is constructed and used, before briefly discussing the normative and empirical implications of the fact that the first phase of the process is one of causal inference.

Estimation - Identifying the causal model

What is presented as transforming the observed distribution into a counterfactual distribution is, in fact, a combination of several analytical steps. Roemer and Trannoy (2016) describe it as follows:

> Socioeconomic advantage has to be estimated through parametric and nonparametric estimation techniques, captured by the coefficient of the circumstance variable in a linear model regressing the outcome on a set of circumstances and effort variables. An evaluation of inequality must be concerned with the process that generates it. This leads Fleurbaey and Schokkaert (2009) to state, provocatively, that any EOp empirical analysis must be preceded by an estimation phase to discover the best structural model leading to the results.

It should be noted that the structural or linear model referred to above is the causal model of the 'process that generates' the outcomes. This is what gives equality of opportunity a substantive interpretation.

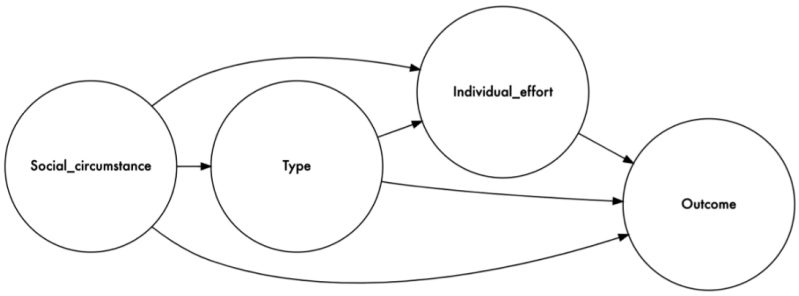
The starting point of indirect methodologies is a causal model of *n* determinants of outcomes, in which all causes of outcomes are classified as circumstances or effort. Here effort is the sphere of responsibility characteristics, where an individual is held responsible for her control or preferences. As noted above, empirical work has predominantly adopted the control view of responsibility, holding individuals responsible for the causes of outcomes they effectively control. This causal model corresponds to a structural model of the system, and thus provides the structure of the structural equation that is used to estimate the structural model from observational data.

The fundamental model of equality of opportunity is presented in ***Diagram 1a***. Here the two classes of causes of outcomes – circumstances and individual effort – are presented as two causes. As we can see, circumstances affect outcomes directly, but also through affecting individual effort. One path of the effect of circumstances is mediated by type, the type being a set of circumstance criteria or a population of individuals fulfilling that set of criteria.

[5] symbols for distributions omitted from the quote

[6] Or more elaborately econometric-estimation stage and inequality-measurement stage.

Diagram 1a: Causes of outcomes



It should be noted that type is an analytic category, defined by selected indicators of circumstances. Later we will come back to the criteria for defining types, but for now it should be enough to note that type is not, strictly speaking, *caused* by circumstances so much as *defined* by them.

Where does the causality come from?

In most analyses of equality of opportunity, the structure of the structural model is not driven by the data but rather based on a prior theoretical understanding of the causalities involved. Data is not used to identify the different causes of outcomes, but to estimate the weight of different causes that have already been identified. This causal understanding, embedded in the structural model used, allows the construction of a model of the determination of outcomes in the system, which allows the construction of a counterfactual distribution by changing the relevant parameters of the model. This is the phase in which prior understanding of the phenomena at play, like intergenerational mobility, is combined with observed interdependencies in the data.

The obvious question is whether this course of action is justified or not. Can we take the model like the one presented in ***Diagram 1a*** and interpret the structural equation describing it as a causal graph? Here it is useful to first take a look at the treatment of causality in the empirical literature on equality of opportunity, and then turn to the causal inference literature to make a sound interpretation of how causality should be understood in the context of equality of opportunity.

Hidden search for causality

There is an extensive and well-developed literature on how a counterfactual distribution is to be produced to measure inequality of opportunity. Despite the fact that Fleurbaey and Schokkaert have explicitly recognised that any empirical analysis of equality of opportunity 'must be preceded by an estimation phase to discover the best structural model leading to the results', the literature seems to somewhat underappreciate or understate the nature of the process as one involving identification of the causal model that produces the observed distribution.

All attempts to measure equality of opportunity frame responsibility in terms of individual responsibility for individual choices or other actions. Roemer (1998) explicitly follows Scanlon (1986) in holding an individual 'morally responsible for behaviour if she decided to do it in a situation where

she was of sound mind'. Ferreira and Gignoux (2008, 2013) present it as the dominant view among social justice theorists that 'differences in achievement that do not reflect the choices or actions' are unjust. Fleurbaey (2008) refers to both Cohen (2011) and Arneson (1990), both of whom attach responsibility to individual choices. When referring to the relationship of circumstances and effort with outcomes, the language of choice usually remains causal and refers to outcomes being determined by effort or circumstances.

But the language of causality is mainly absent from presentations of the methodology of measuring equality of opportunity. Most of the literature adopts explicitly causal conceptions of responsibility but glosses over causality when it comes to the presentation of methodology. The enterprise of identifying just or unjust causes of outcomes – an exercise in causal inference if ever there was one – is presented as focusing on statistical operations. From valuing causes, we come to observing 'joint distributions of circumstances, effort and outcomes', 'regressing the outcome on a set of circumstances' and language on smoothing the distribution when producing the counterfactual distribution. (Ferreira & Peragine, 2015; Temkin, 2013) This is unfortunate, as

> Measuring 'unfair' inequalities implies that a distinction is introduced between causal variables leading to ethically legitimate inequalities and causal variables leading to ethically illegitimate inequalities. (Fleurbaey & Schokkaert, 2009)

Decomposition of variation attempts to construct a statistical model underlying the observed distribution, but it only becomes normatively relevant – and thus relevant as the measure of the construct that is measured, i.e. the effect of circumstances on outcomes – when that statistical model is given a causal interpretation.

Attributing outcomes to circumstances or effort is nothing less than the construction of a causal model that predicts outcomes with two principal classes of causes, circumstances and effort. What is referred to as smoothing the distribution is the presentation of expected values given circumstances, or of counterfactual outcomes given treatment. It removes, in fact, the variation of outcomes due to just causes.

Explicit recognition that the first phase amounts to identification of a causal model before the counterfactuals are defined by the identified model would clarify the issues both normatively and empirically. Normatively, equality of opportunity condemns inequalities caused by factors beyond individual choice or control while holding individuals responsible for the consequences of their actions. This amounts to agreeing with Nozick (1974) that the justice of a distribution 'depends upon how it came about.' Indeed, the difference between substantive equality of opportunity and entitlement theory is not that one measures distributions of states and the other of actions or causes, but what causes are identified as unjust.

Problems of causality and observation

As noted above, while Roemer and Trannoy (2016) 'agree in principle' with Fleurbaey and Schokkaert that 'empirical analysis must be preceded by an estimation phase to discover the best structural model leading to results', they take the statement to be provocative and offer significant reservations on producing a causal interpretation of the results.

Their main reservations are instructive of the challenge of measuring inequality of opportunity. They note that 'reverse causality can often be dismissed, since circumstances are frequently characteristics of states that existed in the past (e.g., one's parents' education)', and are more concerned by omitted variables. They note that endogeneity 'cannot be discarded' because the measurement of equality of opportunity is 'plagued with informational problems', and many potentially important variables, such as genetic variables, are not to be found in the data.

The challenge that Roemer and Trannoy (2016) sketch is the well-known challenge of causal inference from observational data. The downplaying of the causal inference aspect of measuring inequality of opportunity may thus be due to the general methodological reservations about inferring causal information from non-experimental data. (Pearl, 2009; Spirtes, Glymour, & Scheines, 2000)

This cautious approach is, in some respects, eminently justified. After all, measured circumstances are not well-defined interventions (Hernán & Robins, 2020) and suffer from omitted variables, as many authors recognise. (Ferreira & Gignoux, 2013; Roemer & Trannoy, 2016) Roemer and Trannoy note that

> Experiments and quasi-experiments enable one to make causal statements, but experiments can usually only study problems that are much more circumscribed than those that interest researchers in this field. We are trying to understand the whole process by which someone reaches an income level, a health status, or an educational attainment.

Their conclusion is stark and justified:

> These processes are dynamic and cover part of the lifespan of an individual, and understanding them fully in a causal way seems out of reach at present. (Roemer & Trannoy, 2016)

Roemer and Trannoy recognise that the problem of lack of causal interpretation is cause for concern, because 'if we want to give advice to policymakers about the true effect of levelling-the-playing-field policies, impact evaluation needs to be causal.' This warning that, due to omitted variables, the coefficients cannot be interpreted to be causal is very much relevant when we aspire to identify mechanisms of intergenerational transmission of positions and possible policy interventions to affect inequality of opportunity.

But they are taken off course when they give an example to illustrate the question. They ask us to

> consider the case where there is a positive correlation between the health of children and parents. Many different features can explain such a link. Genes, lifestyles, access to medical care, housing conditions (such as the presence of lead in walls or paints), are just examples that come to mind. It is obvious that the remedy, if any, is specific to each case. Whatever the cause, the correlation provides some empirical evidence of violating equality of opportunity. (Roemer & Trannoy, 2016)

What they fail to state is that the correlation only provides empirical evidence of a violation of equality of opportunity *if it is given a causal interpretation* because equality of opportunity is a causal concept. Thus, their position seems to leave them in a bind. It shows that the concept of equality of opportunity requires causality, but a causal understanding of the process is currently[7] out of reach, as

[7] And for the foreseeable future.

is experimental data. Thus, empirical measurement of equality of opportunity must navigate between the Scylla of unwarranted causal inference and the Charybdis of the absence of construct validity. Construct validity requires causality, but data does not readily reveal it. How should these waters be navigated?

2.2 Well- and ill-defined causalities

Well-defined handles

Here, we suggest a solution to the challenge. Obviously, it will not grant us access to universal data sets that cover all imaginable variables and thus remove any doubt regarding unobserved confounders. What it seeks is the conceptual clarity that would allow us to retain the causal nature of the concept of equality of opportunity but to square that with the distinct empirical challenges of making causal inferences from observational data, with large numbers of crucial unobservables.

We may approach the question through the concept of a *well-defined intervention*. Hernán and Taubman (2008) use obesity as an example, arguing that we may not identify the causal effect of body mass index on mortality by comparing different BMIs because no method of intervention on BMI is specified and different methods of intervention may lead to different counterfactual mortalities. They illustrate the conundrum nicely:

> By presenting a single observational estimate for the effect of obesity, the data analyst is either (1) assuming that all interventions on BMI have the same effect on mortality or (2) implicitly considering a complex intervention that changes the determinants of BMI in proportion to the distribution of those determinants in those who are not obese in the population. Unfortunately, both of these may lead to problems. The assumption (that all interventions on BMI have the same effects on mortality) is unlikely to hold true. The implicit intervention (on all the determinants of BMI) is likely to be too difficult to implement to be meaningful to policy makers.

We may note that the concern about the lack of a well-defined intervention is related to the difficulty, or rather near impossibility, of defining and implementing an intervention that would achieve the estimated causal effect. This is equivalent to the concern that 'if we want to give advice to policy-makers about the true effect of levelling-the-playing-field policies, impact evaluation needs to be causal.' (Roemer & Trannoy, 2016) It approaches causality from the same perspective as agency theory of causation (Beebee, Hitchcock, & Price, 2017), taking causes as our handles on the world, places where we can make a difference. A well-defined intervention is one where we can take the handle and turn it to effect change in the world.

Requirements of causal inference

The sufficient conditions for identifying causal effects are positivity, consistency and exchangeability. Positivity requires that all values of the treatment are possible, i.e. they occur. In a randomised assignment, a positive probability for all subjects to be treated can be assured, but without random assignment, the fact that a treatment does not occur in one stratum, or apply equally to everyone,

affects the assessment of the causal effect of the treatment. Consistency requires that the treatment has a consistent effect across individuals. Exchangeability requires that the observed treated and the non-treated are exchangeable, i.e. there is no systematic selection into treatment or missingness of data. Together these conditions ensure that a true causal effect of a treatment is identified, rather than a spurious association of outcome differences with treatment.

These requirements for causal inference can be met to varying degrees and in various ways. While all the requirements are necessary, fulfilling the requirements can rarely be shown and must be assumed to some degree. That is why positivity, consistency and exchangeability are often referred to as sufficient assumptions for causal inference. (Cole & Frangakis, 2009)

The issue is closely linked to the discussion of the role of randomised controlled trials (RCTs) as the gold standard, the ideal method, or even the only reliable method for drawing causal inferences. Their use has spread from medicine to other fields of study, as well as being linked to the 'credibility revolution' in economics.

Randomised controlled trials (RCT) contribute to meeting the requirements of causal inference because both the positivity and exchangeability conditions can be ensured if the allocation of subjects is random, compliance with treatment is ensured, there is no information censoring, and so forth. Quasi-experimental designs achieve both conditions by using settings where both assumptions plausibly hold. The consistency requirement is achieved in RCT through the administration of the treatment by the experimenters; in quasi-experimental settings, it has to be assumed. The plausibility of the assumptions depends on several factors in both experimental and non-experimental settings. (Steiner, Kim, Hall, & Su, 2017)

This is where the need for well-defined interventions comes in. If the treatment is not administered by the administrator of a trial, uncertainty arises as to what degree the indicator of treatment measures a consistent intervention across cases. Ensuring such consistency requires precision in definition and measurement of the intervention – a well-defined intervention. This is the framework in which we need well-defined interventions when we measure causal effects, so that the estimated causal effects can then be used as estimates of effects of policy changes. With well-defined interventions, we are already very close to understanding the mechanism of the causal effect. The better defined the intervention, and the more rigorously the causal effect of the treatment has been identified, the better our ability to estimate in advance the effect of the intervention when it is transferred to another setting.

But the idea that RCTs dominate all other methods of causal inference has also come under criticism. A brief reference was made above to seminal causal inference literature lamenting the general methodological reservations about inferring causal information from non-experimental data (Pearl, 2009; Spirtes, Glymour, & Scheines, 2000). The basics of the criticism have been on the same general lines from the beginning, emphasising that experimental settings operate under the same principles as observational studies. Their relevance for our quest here comes from the fact that the other side of the RCT discussion has been the question of the limits of drawing valid causal inferences from observational data.

> From our perspective outside the discipline, the most urgent questions about the application of statistics to such ends concern the conditions under which causal inferences and predictions of the effects of manipulations can and cannot reliably be made, and the most urgent need is a principled, rigorous theory with which to address these problems. To judge from the testimony of their books, a good many

> statisticians think any such theory is impossible. We think the common arguments against the possibility of inferring causes from statistics outside of experimental trials are unsound, and radical separations of the principles of experimental and observational study designs are unwise. Experimental and observational design may not always permit the same inferences, but they are subject to uniform principles. (Spirtes, Glymour, & Scheines, 2000)

Despite the prevalence of RCTs in medicine, there is also longstanding criticism, in medicine, of over-relying on them and of the assumption that only experimental settings allow causal inference. (Frieden, 2017) Such criticisms have also lately been presented in economics. (Deaton & Cartwright, 2018) The main criticisms relevant to our purposes are related to the transportability of the findings and to the nature of the questions that can be asked and answered.

It has been noted that the conditions of a well-conducted RCT also rarely apply in practice, so that RCTs are also confronted by the same challenges as observational studies. (Hernan, Hernandez-Diaz, & Robins, 2013) RCTs can (if they fulfil all the requirements) provide an unbiased estimate of the causal effect in the study population. But they immediately face the same problems as other methods when it comes to transporting the result to other contexts. This is due to the uncertainties related to the presence of the same causal mechanisms in different circumstances. The estimated causal effect can be transported under two conditions: 'two components must remain invariant: the laws (or mechanisms) and the boundary conditions.' (Pearl, 2009) While an RCT might in ideal circumstances achieve an unbiased estimate in one set of boundary conditions, transportability requires understanding of the effect of those boundary conditions as well. This is more easily said than done, and effectively, in exportability 'there is no serious alternative to trying.' (Deaton & Cartwright, 2018)

Another general criticism mirrors the distinction made by Roemer and Trannoy between answering causal questions and measuring equality of opportunity. Finding well-defined handles is obviously more straightforward in more constrained settings and in answering more circumscribed problems. Well-defined intervention criteria are more readily met by more narrow constructs than broad ones (Holland, 1986) and by down-stream criteria than up-stream criteria (Glass, Goodman, Hernán, & Samet, 2013). These are also the types of interventions that can be more readily researched in a randomised experiment. And of course, in 'the strictest sense, any factor that has human agency is not well-defined'. (Schwartz, Prins, Campbell, & Gatto, 2016)

Questions of the overall effects of social structures on the inequality of opportunity are almost ideally unsuited to be explored by a randomised experiment or in the framework of well-defined interventions. Achieving such definition is currently an unattainable requirement for a field trying to understand 'the whole process by which someone reaches an income level, a health status, or an educational attainment.' (Roemer & Trannoy, 2016)

Well- and ill-defined treatments

But Hernán and Taubman (2008) also refer to another possibility regarding an intervention like changing BMI. We may interpret the hypothetical intervention as a complex intervention, consisting of whatever makes a difference between the treated and untreated groups. In well-defined interventions, we need to understand what the intervention is, i.e. we need to understand the causally pertinent aspects of the intervention, thus approaching an understanding of the mechanism by which the intervention achieves the outcomes it achieves. In an *ill-defined intervention,* the causally relevant aspects

of the treatment may be very much unclear. Indeed, that is what makes it difficult to turn it into an implementable policy intervention.

Social circumstances, such as parental education, are very much like obesity, used as an example by Hernán and Taubman, in that they are far from being well-defined. In practice, in measuring equality of opportunity, the type is accepted as a causal variable as such an ill-defined treatment. That is what justifies the name of equality of opportunity, and the references to the effects of circumstances or of unfair causes on the outcome distribution.

Thus, to be faithful both to the goals of measuring equality of opportunity, the causal nature of the concept of opportunity and to the difficulty of drawing causal inferences from observational data, we should more explicitly understand the estimation phase as causal inference using ill-defined, complex or compound treatments.

This would also help us to recognise that the circumstances we deal with when we assess inequality of opportunity are not meant to be understood as causal handles. We need to package the complex interventions that circumstances are, in large non-transparent boxes, the contents of which are hazy to us. It may say 'parental income', 'socioeconomic status' or 'education' on each box, but these causes should be taken as proxies of causes, not as well-defined causes. Parental education is an ill-defined treatment that is measured by the proxy of formal parental educational attainment. The name on the box is the name of the proxy, not a complete description of a well-defined treatment. It may be thought of as an ill-defined hypothetical treatment that – applying Hérnan and Taubman – changes the determinants of outcomes in proportion to the distribution of those determinants in the untreated group.

It is good to recognise that well- and ill-defined causation answer the causal question on different levels. Thus, the causal effect of parental qualifications on estimated outcomes does not produce an estimate of the well-defined causal effect of giving a parent a piece of paper awarding her a qualification. Even in well-defined settings, qualification is taken to stand for the process that led to awarding of the qualification. In an ill-defined setting, we capture more of the processes that lead to different outcomes depending on parental education. Indeed, by definition, the causal effect of turning a person from one with background a to one with background b, in all causally relevant respects, changes the outcomes of the person from those of a to those of b. Ill-defined causality tries to measure the effect of that implicit, ill-defined complex intervention. It should, and often does, profess to ignorance of the nature of the exact mechanisms involved and of what such a complex intervention would consist of. It may be understood as, if not the limiting case, falling towards the unspecific end of the continuum of specificity of treatments. (Kaufman, 2019)

Because it 'is likely to be too difficult to implement to be meaningful to policymakers' (Hernán & Taubman, 2008) we need well-defined causal inquiries that try to identify the causal handles that can be used to intervene and manipulated to produce different policy outcomes. They are needed to understand what the implicit intervention consists of, and also to see if it is implementable or not. Here, all causally relevant causal confounders must be identified, because otherwise the causal model is misspecified and the estimated causal model misestimated.[8]

When estimating the well-defined causal effects of the mechanisms where the causes are handles that allow manipulation, it is essential that the model includes all causally pertinent causes, or

[8] In Chapter 9: Genetics of sex, race and class, we will return to the analogous relationship between the weak and strong biological explanations.

the causal effect will be misspecified. When measuring ill-defined upstream causes as when measuring equality of opportunity, the situation is different because indicators are not taken to capture well-defined handles. Thus, while the causal model of a well-defined intervention contains uncertainty as to the identification of all causal confounders (which can be understood as well-defined causes missing from the graph) in an ill-defined model the uncertainty is, to a higher degree than for a well-defined one, expressed as uncertainty of the content of the treatment. There is always uncertainty as to the causally relevant content of the intervention, but with ill-defined causes that uncertainty is obviously much greater.

Embracing ill-defined causality

Embracing ill-defined causality is defended here as a means to clarify the content of equality of opportunity and the project of empirically identifying it.

First, the question of causality is unavoidable in this context. The only alternative to embracing ill-defined causality is to simply hide any complexity under the rug and wait for our understanding of the process on the level of well-defined treatments to improve considerably. Conceptually, responsibility requires choice, which brings in causality and counterfactuals, so throwing causal inference out is not sustainable. The fact that causality becomes relevant through the questions we want to answer has been beautifully argued by Hérnan (2018) in the context of public health:

> The lack of clarity regarding the goals of the research has often been justified by the questionable validity of causal inferences from observational data. However, this argument simply conflates the aims and the methods of scientific research. An association measure from an observational analysis may be a biased estimate of a causal effect but being explicit about the goal of the analysis is a prerequisite for good science. Do we want to estimate the association measure or the causal effect measure?

At the same time, it is imperative to recognise the limitations of such exercises. The cautious solution advocated here is to combine causal interpretation with an upfront recognition of the assumptions and limitations of the estimation of the causal effect.

More explicit acceptance of the use of ill-defined treatments and their difference from well-defined ones might also help make sense of the example of Roemer and Trannoy (2016) of the health effects of lead paint on parental and offspring health. They are obviously right that poor parental health is not a well-defined treatment that causes poor offspring health when that cause is the presence of lead at home. But if we think of parental health as an ill-defined treatment, shades of grey arise. If the measure of parental health even indirectly involves lead as a cause of poor health, the health measure becomes a proxy for a relatively poor socioeconomic status, as exposure to lead is not socioeconomically random. It may be a useful proxy for defining a type, even though it is not a well-defined treatment.

This position seems close to the one expressed by Roemer and Trannoy (2016). While we do not yet understand the mechanisms of intergenerational transmission of positions, income or educational attainment well enough to design *well-defined interventions* that would change the outcomes, we may obtain useful results when we interpret them as *ill-defined interventions*. In this approach, we admit our ignorance of the exact mechanisms and content of the treatment, leaving inquiry into which to methods more suitable for the purpose. Ill-defined measurement of equality of opportunity

may still produce valuable knowledge about the differences between systems – or change over time within a particular system – of the level of equality of opportunity. Focusing on those changes, and their co-occurrence with measured phenomena conceivably acting as causes, or following as their consequences, may provide us clues to where to search for well-defined causal effects.

In policy terms, it may be useful as well, even if it does not provide well-defined causal understanding. Recent experimental work has demonstrated that, at least in the improvement of physical tools, understanding of the causal mechanism of their function is not necessary, in the presence of cultural learning. Indeed, significant improvement has been achieved in practical performance with no improvement of the causal understanding of the process. (Derex, Bonnefon, Boyd, & Mesoudi, 2019) The fact that such improvement does not require a causal or full understanding of everything relevant at play has of course been emphasised by numerous writers since Mandeville (1989).

Summary - Model from the machine

All causal inference requires assumptions. As noted above, direct control over the research setting helps to make those assumptions more easily warranted, even though they can never be completely fulfilled. In observational research, the underlying assumptions can be expressed as a causal model that describes the causal structure at play.

The standard strategy is to base the causal model on a prior theoretical understanding of the causalities involved. This pre-existing theoretical understanding can take the form of a causal diagram like the one in ***Diagram 1a*** above. The diagram is simplified in the sense that it does not differentiate between different circumstances or different forms of effort. The questions on the number and choice of indicators of circumstances shall be discussed below. In this phase it should be recognised that, in the standard empirical approaches to equality of opportunity, data is not standardly used to identify the different causes of outcomes but to estimate the weight of different, already identified causes. This causal understanding, embedded in the structural model used, allows the construction of a model of the determination of outcomes in the system, which allows the construction of a counterfactual distribution by changing the relevant parameters of the model. In empirical research on equality of opportunity, the causal model includes a relatively small set of unjust causes of outcomes, such as gender or socioeconomic status. These unjust causes of outcomes are most often implicitly embedded in the choice of indicators for the specification of type and amount to specifying an ill-defined causal model even when it is not explicitly conceived as such.

Such a situation is obviously vulnerable to misspecification of the causal model as a model of well-defined interventions, and the validity of the predictions of the model are conditional on the model being correctly identified. We may be confident that any model now in use is misspecified in the sense that giving individuals the minimum required to fulfil the criteria of causes used in the model, a piece of paper indicating the award of a formal qualification, would not have the effect that the estimation produces. The indicators of causes that are standardly used capture the effects of causes that lead to those observed indicators of causes, and also to the observed outcomes. We are, in crucial respects, utterly ignorant of behaviour of the more precise mechanism of these effects.

3 Types, effects and distributions of opportunity

3.1 Type and treatment

Ex-ante and ex-post

Measuring inequality of opportunity is approached from ex-ante and ex-post perspectives that differ on the amount of information available. From the *ex-ante perspective,* opportunities are assessed in terms of the value of opportunity sets available to individuals in their circumstances. Circumstances are the basis for dividing the population into types that share the same circumstances. In the *ex-post perspective,* the effort is determined as well and forms the basis for dividing the population into tranches that are defined by effort. The ex-ante perspective thus tries to identify unjust inequality that is due to circumstances, while the ex-post perspective assesses how outcomes correspond to effort.

In the ex-ante approach, the population is divided into types that share the same circumstances. The ex-post approach divides the population into tranches that share the same level of effort. The distribution of outcomes is then assessed against the background of types or tranches, and both strategies involve the division inequality of outcomes into between-types/within-tranches inequality and within-types/between-tranches inequality. In the morally meritocratic equality of opportunity approach, the principle of horizontal equity is present in the demand that individuals with the same level of merit should be treated equally, and the principle of vertical equity is present in the demand that individuals with different levels of merit should be treated differently. Between-types within-tranche inequality is unfair because it reflects circumstances rather than merit. Within-type between-tranches inequality is just and corresponds to rewards for effort. In the ex-ante strategies, inequality of opportunity decreases if inequality between individuals with different circumstances decreases. In ex-post strategies, inequality of opportunity decreases if outcome inequality decreases among individuals at the same degree of effort. (Checchi, Peragine, & Serlenga, 2010; Ferreira & Peragine, 2015; Fleurbaey & Peragine, 2013)

The ex-ante approaches can be demonstrated by ***Diagram 1***, which illustrates how key features of circumstances are encapsulated by type. Because circumstances fully define the type – and effort is independent of circumstances – there is no causal path from responsibility characteristics to type. Within each type, individuals differ in

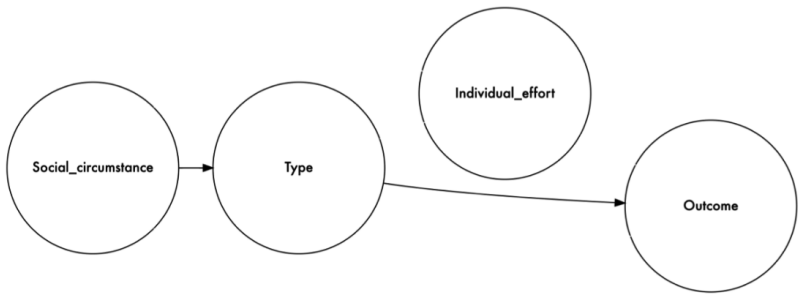


Diagram 1b: Causes of expected values of outcomes given type

their circumstances and their responsibility characteristics, which is denoted by the direct causal paths from circumstances and effort to the outcome (***Diagram 1a***). Taking only the between-types variation of outcomes eliminates the direct causal paths from circumstances and effort to outcomes, i.e. all differences in circumstances and effort that are not mediated by type ***(Diagram 1b***). This, of course, amounts to eliminating all within-type variation outcomes from any counterfactual distribution produced by the model.

In both approaches, what is identified is the effect of causes in a causal model structure of which is based on a preliminary understanding of the factors at play. If the presupposed causal model is correct, what is identified is the causal effect of circumstances on the distribution of outcomes, as all the effects of effort have been removed. (Pearl, 2009)

Type as treatment

The step from observation of associations between background characteristics and outcomes to identifying unjust *causes* of outcomes relies on taking type to be a treatment. The effect of unjust causes on outcomes is inferred from the observed distribution of outcomes relative to type. Defining types assumes that outcomes varying *with* background characteristics vary *due* to background characteristics. This causal assumption moves us from assessing the distribution of outcomes – or even the covariation of outcomes with circumstances – to assessing the effect of circumstances on outcomes.

This is the crucial assumption of standard methods for measuring equality of opportunity. Types are interpreted to be treatments and treatments are identified by the background characteristics of individuals. When inequality of outcomes is assessed against the background of gender or parental income, it is interpreted that, an intervention on an individual's gender or parental income, with all the changes in treatment that entails, would have had a corresponding effect on the outcomes of the individual. This has some important implications.

The embedded causal assumptions distinguish measures of inequality of opportunity from those of outcomes. While equality of opportunity is measured by comparing groups, the difference between the groups is taken to be a difference caused by a difference in treatment. As noted above, this is by definition true when we take the treatment to entail everything that makes type A different from type B. The treatment may be ill-defined, so that we do not have an inkling what the effective treatment consists of, but type still matches a hypothetical treatment and the between-type variation measures causes, rather than outcomes.

Understanding of type as a treatment also shows how critiques that take such approaches as collectivist rather than individualistic (Lippert-Rasmussen, 2011) miss the mark. The treatment does not take types or groups as functional entities, but as being defined by parameters of the treatment they have received. Inequality of opportunity is indifferent to who is responsible for the treatment that leads different types to have different outcomes. It identifies differences in the expected outcomes of types as unjust even if the expected outcomes would differ only because of how individuals of the same type treat each other. The treatment would be unjust even if all the members of the group were both offenders and sufferers of the unfair treatment. This perspective is profoundly individualistic, rather than collectivist.

Number of types

When defining types, the number of circumstances to be taken into account quickly increases the number of types. Typologies used in research range from measuring circumstances with one non-granular indicator to extensive typologies with many type-defining indicators and correspondingly higher numbers of types. (Björklund, Jäntti, & Solon, 2007; Smeeding, Erikson, & Jäntti, 2011) Usually it is necessary 'to choose a fairly small number of circumstances, each of which can take on a fairly small number of values.' (Roemer, 2013)

The fewer indicators used, the less able the analysis is to observe different aspects of inequality or inequalities that emerge when several types of inequalities are compounded, and the more the estimate of inequality is a lower bound estimate. What is gained from a smaller number of indicators is a less elaborate and unwieldy system of types, as well as a possibly more reliable estimate of observed levels of outcomes per type. In general, types are defined by a limited set of variables, thus necessarily sacrificing specificity in the causal effects identified. The effects identified are not on the level of mechanisms, and thus types are not well-defined treatments (Hernán & Robins, 2020), even if they should be understood as treatments.

Some lack of definition of the treatment is of course precisely what is to be expected from the indirect method as it tries to capture unjust inequality even while ignorant of both the theory of exact mechanisms of treatment and of data to directly observe those mechanisms in action. The power of the results depends on how justified the initial assumptions on the causal model are.

Initial circumstances and equality of opportunity

Partly due to limitations of available data on circumstances over a lifetime, much of the empirical research on equality of opportunity has focused on initial circumstances, i.e. background. Circumstances long after birth come into play to the degree that they are associated with the background circumstances used to define and identify types. Thus, luck that systematically favours or disfavours a type is captured, because the focus is on the consequences of type.

Initial circumstances are usually measured by ethnicity or some measure of socioeconomic status (SES). Different measures of socioeconomic status have been divided into cultural, social and economic capital. This division is close to the three most common individual measures of SES: education, occupation and income. The measures capture different aspects of SES and are not interchangeable in the analyses. Education and cultural capital focus on cognitive and non-cognitive traits, social capital on social status and income on material resources. (Bukodi & Goldthorpe, 2012; Jæger, 2007) Measures of cultural capital are usually aggregates that include measures of formal education, measures of social capital measure social status and networks, while measures of economic capital are primarily built around income. Measures of cultural capital also often include more indirect measures, such as measures of participation in or attitude towards high-brow cultural activities or number of books at home. (Kraaykamp & Van Eijck, 2010; Schütz, Ursprung, & Wößmann, 2008)

Education, occupation and income, as well as cultural, social and economic capital, are strongly associated. Education leads to access to occupations, occupations lead to levels of income, and all of them contribute to measures of cultural capital such as participation in high-brow cultural activities, number of books at home or subscription to newspapers. Consequently, many of the effects of education, occupation and income are shared. Of the three, parental education (Bukodi & Goldthorpe,

2012) and cultural capital (Møllegaard & Jæger, 2015) seem to contribute more than income and occupation or economic and social capital on the outcomes.

Parental education is one of the most widely used indicators of circumstances in the study of intergenerational mobility and equality of opportunity. It is used as a standard example of defining types used in a range of studies, from those where it is the only indicator used to define types to being one criterion among many in multidimensional typologies (Roemer & Trannoy, 2015; Björklund, Jäntti, & Roemer, 2012). Referring to Roemer, Corak (2004) contends that

> parental education [...] may offer a more appropriate indicator of the influence parents may have on the aspirations and personality of their children.

In research on learning outcomes in youth, the problem of a high number of missing observations of parental education in, e.g., the large-scale survey Trends in International Mathematics and Science Study (TIMSS) has led to the use of books-at-home as a measure of SES instead of parental education. The choice has been supported by results showing that, at least in some cases, it is a more significant predictor of student performance than parental education. It has been interpreted as a proxy for a family background that values education and academic success, but also a proxy for family economic background. (Schütz, Ursprung, & Wößmann, 2008) The challenge is that as the numbers of books at home are self-reported, and in adult surveys, the number of reported books at home at age 16 may, for a large proportion of respondents, be reported after a very considerable time lag. Thus, there may be a considerable risk of reverse causation if more academically oriented adults remember higher numbers of books than those of their peers who are uninterested in books.

3.2 The positional and the absolute

Positional and non-positional outcomes

Understanding explicitly that type is used as a proxy for treatment may clarify what indicators of circumstances we use to define types. In essence, we want the type to be defined by circumstance criteria that imply the same causal treatment in all compared cases so that the exchangeability assumption is fulfilled. Thus, the definition of type is tied to the questions asked and to an understanding of the causal significance of different aspects of observed circumstances.[9]

First, we must consider the outcome. We may divide outcomes into positional and non-positional or absolute goods. Positionality means that the value of a good is affected by the consumption of others. The intrinsic value of a good does not decline when it becomes more abundant. Positional goods, which derive value from relative scarcity, decline in value when they become more prevalent. An axe is a non-positional good to the degree that it directly brings benefits via chopping wood, and positional to the degree that axes are status objects reserved for the elite and thus offer different types of economic, social or political benefits. Non-positional goods are measured by the level of holdings

[9] Here, some challenges are created by the fact that circumstance criteria used in an observational setting are almost invariably and inevitably so non-specific that they cannot be considered well-defined interventions. Thus, the causal significance of different aspects of circumstances should be understood as a question of the principal nature of the well-defined treatments that form the complex composite treatment proxied by circumstance criteria.

in income, education, or axes *on an absolute scale.* Positional goods are measured by the relative position in the ranking of those holdings *on a positional scale*.

The significance of the nature of the supposed outcome is relevant because any effect of the cause on the outcome depends on the scale of measurement of the outcome. If we measure axe-ownership in non-positional terms, increases in axe-holdings can be associated with more chopped wood. In positional terms, axe-holdings cannot improve in the aggregate, because a position in the queueing model of axe-holdings does not depend on the number of axes held. In contrast, the status advantages provided by axe-holding are zero-sum and thus become inflated when holdings increase.

In short, positional goods are zero-sum and their effect depends on their distribution, while non-positional are not zero-sum and their effect is independent of distribution. The positional nature of goods may be taken into account in two different ways. We may simply measure the relative scarcity of the good so that it becomes less valuable as it becomes increasingly common. Another way is to measure the positional value of a good in terms of another good. Here, obviously, the choice of the other good is crucial, because it amounts to a selection of the currency in which the value of a good to be determined and the results may depend on the currency. Relative values of educational qualifications may change differently over time, depending on whether we measure the value of the qualification by additional skills gained, additional income earned or by ranks climbed in the social hierarchy. As Schneider (2016) notes in the case of education:

> Moreover, a scaled education variable is no longer a pure education variable because it expresses education in the 'currency' of another, quantitative, variable, as it were. Thus, the comparability problem shifts to this other variable and, in the case of differences over time or across countries, it is unclear whether they are due to changes in education, to the scaling variable, or the relationship between the two.

There are obviously differences in the fixity of scales, depending on the positionality of the scale of measurement. We may imagine pure status goods that are only valuable in terms of their relative scarcity and the relative rank of an individual in the holdings ranking of the status good. For such a good, the difference between the 10^{th} and 90^{th} centiles (or between the 25^{th} and 75^{th} centiles) is functionally the same irrespective of the difference between the quantiles in terms of the absolute values of the status good held, and the scale is completely fixed in the sense that the aggregate value of relative holdings cannot change if the value is purely positional. That, of course, is a limiting case, describing pure positional good that only has value by virtue of its scarcity. The limiting case, at the other extreme, is just as obviously one where positional holdings play no role in value of the good.

Positional and non-positional treatments

The nature of the relevant outcome has significant implications for our analysis. The choice of the currency of the outcome defines the currency of the effect that we may identify. When we measure outcomes in positional terms, we may only identify positional effects. Thus, when education or income is assessed as a positional good, the burden of comparability over time shifts to the comparability of positional ranks over time.

This may have been underemphasized in recent sociological literature, which has increasingly turned towards positional measures and estimated e.g. the *social class returns* to education. In contrast to economic returns to education, which are rising even though educational attainment is rising, social class returns are declining. (Bukodi & Goldthorpe, 2011) What has not been emphasised enough in the analysis, perhaps, is that class is a positional concept that does not imply any one set of outcomes. The working class in England is substantially different from the eponymous class in the mid-1800s in very many important respects. Any rise in education levels cannot make everyone middle class, because middle class conceptually requires something above and something below. Thus, the result of Bukodi and Goldthorpe (2011, 2012, 2016), that the amount of education needed to achieve a given class position has increased, would seem to follow inevitably from a combination of educational expansion in absolute terms and the class position scale, which is more fixed. Such reduction of class returns is not as unavoidable with rising levels of education as a reduction in social rank returns because the class structure may evolve somewhat, even if it may not necessarily evolve so that everyone becomes upper class. Rank measured by quantiles obviously does not change at all, so any rise in the level of a good leading to rank will lead to an inflation of that good as measured by rank returns.

The choice of scale affects what effects of treatments can and cannot be disentangled. A purely positional conceptualisation of treatment assumes that the causally relevant difference is between quantiles in the distribution, and thus that a treatment that changes position from 10th to 90th centile (or between the 25th and 75th centiles) is functionally the same irrespective of the difference between the quantiles in terms of dollars, years in education or axes held. In such a setting, a difference of 10 units in axes or years of education is causally equivalent to a 1-unit difference if both represent a difference of, e.g., 80 percentiles in the social ranking. Where the measure is limited is in not being able to distinguish between the effects coming from the number of axes and the effects of one axe. Thus, if axes do, contrary to assumption, have absolute effects as well, they are invisible in the analysis. Correspondingly, a purely absolute measure would mean that changes in years of education or axe-holdings have the same effects irrespective of the distribution of education or axes. Comparisons over time with absolute numbers of axes do not recognise changes due to a single axe giving an advantage of 10 ranks in the positional queue at one time and only 1 rank in another. When we choose to hold positional or absolute distance constant over time, any effects of the change in the other become invisible.

It may be useful to recognise that no good is strictly positional or non-positional. It is positional in some respects and non-positional in others. Thus, noting that parents or employers 'may themselves view education in relative terms', as noted by Bukodi and Goldthorpe (2016), is not a sufficient cause to measure educational attainment with a positional measure. It is too facile to simply choose absolute or positional measurement by showing that the other option is untenable in an extreme form.

Assuming that formal parental educational attainment is not positional at all amounts to an assumption that 'having a parent with a university degree always confers the same absolute advantage, regardless of whether that parent was among the elite few who earned a degree in the mid-20th century or the larger share who earned a degree at the turn of the 21st century.' (Chmielewski, 2019) This observation is absolutely correct, but it does not amount to a sufficient reason to measure educational attainment in positional terms. It merely observes that an absolute measure of education does not fully capture the effects of education, which are partly positional. If rejecting either full positionality or a fully absolute nature of educational attainment would be sufficient to embrace the

other, we could just as well embrace absolute measurement after noting that assuming full positionality is not credible. Assuming that formal educational attainment is only positional would imply that raising average education level from an average of three years of formal education to 15 would have no real effects on society, or that a 90/10 gap of 1 year of formal education is functionally equivalent to a 90/10 gap of 10 years of formal education to the relative positions of the individuals. Such a position would be difficult to defend.

Neither does positional measurement follow from the facts that 'dollars are not comparable across years, and […] income inequality changes over time'. Contrary to what has been implied, the choice of positional measurement does not provide a 'method that characterizes the relationship between income and children's achievement in a way that is comparable over time' (Reardon, 2011) unless income is a purely positional treatment. Most methodologies that make financial resources comparable across time do not assume money to be only positional and correct for inflation to account for the changing value of money. The observed associations between income inequality and intergenerational mobility, the so-called Great Gatsby Curve made famous by Corak (2013), and the mediating role of education in the process (Cingano, 2014), would also imply that income is not purely positional, and the changing income distance between rank positions has real effects. If income were purely positional, measures such as the Gini coefficient would be devoid of interest as they could be used to measure either something substantively irrelevant – like dollars and euros – or something fixed by definition – the rank order. This does not imply that income should not be measured in positional terms, but it does imply that a stronger theoretical justification might be in order.

Because education is both a positional and a non-positional good, the choice between positional and absolute measures depends on whether education is positional or absolute in the relevant respects. The relevant respects depend on the question at hand. In the end, 'different research questions and hypotheses can be investigated with positional measures than with absolute measures, so that the measurements, too, can meaningfully complement each other.' (Schneider, 2016) While the development of between-group income inequality between education groups is in line with the overall wage inequality development in the US, the questions are quite distinct. (Autor, Katz, & Kearney, 2008) Both choices carry with them underlying assumptions on the nature of the treatment and thus have both strengths and weaknesses.

Education as a non-positional and positional good

The difference between relative and absolute treatments is very much present in the area of education, because of the significant changes in educational attainment over the last two centuries. Tertiary education carried an elite status in many countries relatively recently, and the widening access to tertiary education entails that it is no longer uniquely reserved for the social elite. When using parental education as a criterion of type, the choice to make is between defining parental education by absolute educational attainment – tertiary, below lower secondary education – and relative educational attainment – attainment at 90th or 10th percentile. But which aspect should we take to be more meaningful if we use parental education to define type?

First, some features of education can be considered non-positional. Here it is essential to note that, while outcomes are measured on an absolute scale, our interest does not lie in whether the same cause always leads to the absolutely same result, but whether the same absolute differences in cause lead to the same corresponding differences in outcome. Thus, while the scale is absolute, the quantity

of interest is the reward function, which is relative. Thus, measuring educational attainment using qualification level does not imply that the same qualification should always produce the same outcomes, but that the distribution of outcomes in absolute terms tracks the distribution of educational attainment in absolute rather than positional terms.

Here it may be useful to note that there is substantial variation between countries, and between age groups in countries, in the level of cognitive skills implied by educational achievement. Moreover, there is wide variation in observed skills within each level of education. (OECD, 2013b) But there is much less variation in the difference of outcomes between, e.g., tertiary and upper secondary graduates or between upper secondary and lower secondary graduates. More importantly for comparison across time, when the learning outcomes differences between tertiary, upper secondary and lower secondary graduates are compared across age groups in different PIAAC jurisdictions, it can be observed that the learning outcomes difference is relatively stable across age groups for most countries. This would suggest that the effect on skills of the move from one, roughly measured, level of qualification to another has not changed over time, while the selectivity of each level of education has changed.

Another example may be taken from income. The labour market effects of education are usually interpreted within three different theory frameworks: human capital, job competition and social closure. The human capital theory posits that education produces skills that raise productivity, which is in turn rewarded by the labour market, rendering the wage premia of education returns on skills produced by education (Mincer, 1974; Becker, 2009). In the job competition model (Thurow 1975) and signalling theories that arose as a critique of human capital theory (Arrow 1973), jobs are ranked according to their complexity and demands, while applicants are ranked based on observable characteristics, with formal education among the most important criteria, and the highest-ranked jobs go to the highest-ranked applicants. Finally, social closure, or credentialist hiring, posits that credentials are used to limit access to occupations to create an artificial shortage of labour. Returns on education are a function of negotiations, rather than arising from interpersonal differences in productivity. The rising economic returns on education in, e.g., the US (Autor, Katz, & Kearney, 2008), and in many other countries, have been taken (Pfeffer & Hertel, 2015) to support human capital theory and the skill-biased technical change hypothesis in particular because income returns on education have increased alongside a massive expansion of education that has, by definition, implied credential inflation in terms of the social rank attached to a given education level.

Defining types by educational attainment

In empirical research on the inequality of opportunity, parental education is one of the most widely used measures of background. The scale used is most often absolute and takes on a very limited set of values. While rough classifications into graduate/non-graduate parents are sometimes used, the population is most often divided into three groups: 1) neither parent has attained upper secondary education, 2) at least one parent has attained secondary or post-secondary, but non-tertiary education and 3) at least one parent has attained tertiary education. The three-level absolute scale of education is most often used as one of several measures of background, but it has also been used as an exclusive measure of type. (Björklund et al., 2012; Checchi, Peragine, & Serlenga, 2010, Roemer, 2013) When sample size allows, more granular education classifications are also used. (Checchi & Peragine, 2010)

Another relevant consideration relates to the stability of the classification over time, because the primary-secondary-tertiary division of education classifications has been fairly steady since the UNESCO 1958 Recommendation Concerning the International Standardisation of Educational Statistics was published.

In research on educational inequality, education has been primarily, even almost exclusively, measured using absolute measures. In reaction to this situation, positional measures have recently grown more common, in recognition of the fact that educational attainment has a positional aspect as well. (Schneider, 2016) Research using PIAAC and IALS data by Green and co-authors has measured parental education in absolute terms as a non-positional variable, using the standard three-level classification of education by level of formal qualifications into high, middle and low. (Green, Green, & Pensiero, 2015) In contrast, Chmielewski (2019) uses parental education as a positional variable in her analyses of a comprehensive set of international large-scale surveys of youth skills between 1964 and 2015. For her, the top/bottom types are represented by the $90^{th}/10^{th}$ deciles, following Reardon (2011), who used the same $90^{th}/10^{th}$ deciles classes in his work using a wide variety of data sets.

Implicit or explicit use of parental education as a positional variable is more common when it is used as one of several indicators of parental SES in a composite indicator of SES. For example, Broer, Bai and Fonseca (2019) analyse trends in socioeconomic inequalities in TIMSS and use parental education as one of three variables in a composite SES indicator based on parental education, household possessions and books-at-home. Whether parental education is strictly positional is debatable, because a composite score is calculated from the 'absolute' values of each indicator (education contributing max 4 of the max 10 points of the composite indicator) and the values of the composite indicator are then estimated for the 75^{th} and 25^{th} percentiles.[10]

In terms of the cognitive effects of education, we interpret that education is considerably, but not necessarily exclusively, a non-positional good. Based on the above, it is considered here that, while there may be differences between countries in how large a difference in learning outcomes the level of qualifications translates to, there is a difference, and that difference is mainly on the same level in all age groups, implying that there has been no significant change over time. This is taken as licence to consider that relatively stable differences between groups in cognitive performance are captured by the differences in educational qualifications, even if there is important variation in cognitive outcomes within all educational attainment groups.

In terms of the economic effects of education, it is interpreted that education is largely, if not exclusively, a non-positional good, as suggested by the human capital theory. The human capital theory has been vigorously attacked in the past, the attacks ranging from the insightful (Arrow, 1973) to the vocal (Caplan, 2018). Thus far human capital theory has stood up well, and while it is far from being shown to be ultimately right, the success it has had warrants its use as a framework here as well.

While there are good reasons to consider that education has important aspects of a positional good, this author feels that the considerations presented above warrant measuring parental education

[10] For those interested, it may be of interest to note that the TIMSS classification of education differs from the more common variants as the derived parental education variables that combine post-secondary non-tertiary with short-cycle tertiary (TIMSS 2015) or classify those with some university studies together with those who have upper secondary degrees (TIMSS 1995). This results in a classification that cannot be reduced to the common 3-level classification because the categories used go across the usual classification border between secondary and tertiary education.

on an absolute scale. Such an absolute measure may of course be supplemented with a positional measure, as will be done in the empirical sections of this work.

3.3 Confounding and the relevance of outcome

Variation within types

Variation within types is taken to be normatively unobjectionable, as it is taken to reflect effort. If we define type by parental education, there are always parental income, gender, ethnicity, region and several other factors that may conceivably exercise a causal effect on the outcomes of interest. If there are circumstances that affect outcomes but are not captured in defining types, and we may usually take them to be numerous, 'the inequality of opportunity […] should be considered to be a lower bound on the true inequality of opportunity.' (Roemer & Trannoy, 2015) How much of a lower bound estimate depends on the significance of omitted variables for the outcomes considered.

Even when we measure parental education, for example, the low granularity of the measurement may lead to under-identification of circumstances. Within the standard classification of education into low (lower secondary), middle (upper secondary) and high (tertiary) education, there are significant variations within the classes. University education is often very different from other, non-traditional forms of tertiary education in terms of status, later labour market outcomes, and income. Within universities, fields of study differ, with high-status fields such as law and medicine leading to rather different outcomes from some humanities or arts. Within fields of study, the university attended may play an important role. And so on. All shortcomings in granularity make individuals responsible for their background to the degree that pertinent features of the background go unrecognised.

As Checchi, Peragine and Serlenga (2010) note:

> It is important to remind that by effort in this paper we mean not only the extent to which a person exerts himself, but all the other background traits of the individual that might affect his success, but that are excluded from the list of circumstances. […] This assumption may lead us to overestimate the portion of inequality which is ethically acceptable, and conversely to underestimate equality of opportunity.

That circumstances may affect individual effort is not a problem for the assumption that circumstances and effort are independent. It is necessary to remember that the assumption of independence of circumstances and effort takes effort to refer to the sphere of individual responsibility, and circumstances as the sphere beyond individual responsibility. Thus, they are independent by definition. But circumstances obviously have effects on how individuals behave and on the observable effort they expend. Confusion may arise, as here effort has the more commonplace meaning of a laborious attempt. Roemer marks the difference by referring to measured effort as *raw effort* and contrasting it with morally *accountable effort.* The latter is independent of circumstances, the former is not.

When effort is under individual control, between-types variation in effort can be interpreted as being caused by the incentives faced by the types. Circumstances affect individual incentives, and different groups in different positions might present different distributions of effort because they face

different opportunity sets, and thus different structural incentives. Groups facing much steeper obstacles to entering higher education may, completely rationally, expend less effort to do so. Thus, their lower levels of effort are caused by their circumstances. (Fleurbaey, 2008) When individuals are held responsible for features subject to individual control, socioeconomic status may not act as a mediator of responsibility characteristics, as individuals do not choose their families.

A step towards accountable effort can be taken 'by comparing his behaviour only to others who share his circumstances.' (Roemer and Trannoy, 2015) Roemer accordingly converts measured *raw effort* to morally *accountable effort* by purifying it from the effect of circumstances on effort. This is done by holding individuals responsible for their positional effort, not their absolute effort. In his ex-post approaches, 37th effort percentiles are compared across types, even if the same percentile might have different levels of observed effort in different types. This strategy assumes that between-type differences in the observed effort are a function of the incentives that are faced by different groups and that there are no inherent differences between groups in the amount of effort they are able to expend. (Roemer and Trannoy 2015)[11]

This cut between observed and accountable effort implies the distinction made by Temkin between doing well and being well, referring to the difference of doing good deeds and possessing high moral virtue. He argues for the latter:

> If, in fact, Randi would do everything Alan actually does, if she were in his position, and vice versa, then on the view of proportional justice I favour Alan and Randi ought to fare equally well, from the standpoint of justice, even if in fact, given their different circumstances, Alan actually *acts* rightly, and *does* good, more often than Randi. (Temkin, 2011)

The question of raw and accountable effort is not directly relevant to the empirical analyses in this work because we follow the ex-ante approach that 'views inequality of opportunity as reflected in the degree to which average outcomes of different types differ.' (Roemer & Trannoy, 2016) It is, nonetheless, useful to make a note of it, because it highlights the nested structure that any measurement of equality of opportunity must deal with. Individual choices are limited by social choices and structures, which are in turn limited by what is possible with the technology – understood in the broad sense as all the physical, abstract and social means or tools at our disposal – available to us in the world that we inhabit. The further upstream the cause, the more it affects the outcome distribution indirectly, through long causal paths. And the further upstream the cause, the more difficult it may be to distinguish between what is and what could be.

Confounding unobservables

As noted, the measurement of equality of opportunity using observational data is 'plagued with informational problems' as omitted variables are widespread. (Roemer and Trannoy, 2016) While the challenge of unobservables may not be solved in the foreseeable future, possible confounders may be

[11] There are views that raw effort should be respected, instead of standardising effort differences between groups. This view is based on the idea that, even if the effort differences have been caused by differential environments, the different levels of effort have been actually expended by the individual. (Roemer, 1998) This choice is linked to allowing bare, morally arbitrary facts to have normative implications, and will be directly addressed in the next chapter, which also addresses the broader question of responsibility characteristics beyond individual control.

statistically[12] controlled for. Here we run into some challenges when measuring inequality of opportunity.

The first important fact to note relates to the role of control variables, the choice of which depends on the question to be answered. When the focus is on overall inequality of opportunity by background, many indicators of individual circumstances at the time of outcome measurement may be mediators of the effect of background characteristics rather than confounders, and thus controlling for them would lead to confounded estimates of the effect of background characteristics. When parental education leads to outcomes through multiple channels, controlling for attainment in some of those channels produces a confounded estimate of the effect, because part of the effect is left unobserved. For example, controlling the effect of parental education on learning outcomes by own education, income or occupation would lead to overcontrolling, as the effects mediated by educational attainment, income or occupation, respectively, would be eliminated from the estimand.

There the general principle at play has been known in causal inference for a while:

> Effects are made independent when we condition on their common causes, and effects are made independent of their remote causes when we condition on their more proximate ones. (Spirtes, Glymour, & Scheines, 2000)

Thus, while controlling for many confounders might be appropriate when exploring the mechanisms of causation between background and outcomes, it may not be appropriate when estimating the overall effect of background on outcomes. Any analysis of an upstream cause will find no effects if the downstream causes that mediate the upstream cause are statistically controlled for. This also shows the difference between the approach needed to find the effects of well-defined interventions, as opposed to the effects of multifaceted social conditions. As we are interested in the overall effect of circumstances, and not in the mechanism, controlling for mediators is not in order.

The other set of possible confounding variables precedes background variables, as causes of both type and outcomes. Roemer and Trannoy refer above to genetic variables as an obvious example, but we may simply think of any cause that might both cause type and the outcome. Here, one challenge is data availability, as common causes of background and outcomes may not be easily measurable. Another is the fact that these causes are those well-defined treatments that we are seeking in the long run. At the moment, we do not really know enough about the causalities involved to control for these common causes, even when we can measure them. This is apparent for genetic differences, as will be more explicitly argued later, because we very rarely have strong genetic explanations for the association between genes and cognitive outcomes.

Estimation depends on the outcome to be estimated

Recognising the exercise of measurement of equal opportunity as one causal inference helps us to note that different outcomes of interest – income, welfare, educational attainment, learning outcomes

[12] For the most part, specification statistically is strictly unnecessary in this work, as all settings are observational and thus experimental control of variables is not possible. But the specification acts as a reminder to the author and the reader that statistical control is not equivalent to experimental control.

– may have different underlying causal structures. While the effect of innate ability on learning outcomes may be mostly unavoidable (if we rule out levelling down), it may not be unavoidable for welfare outcomes, insofar as welfare outcomes can be redistributed by e.g. redistribution of income.

Thus, recognising that measuring equality of opportunity is partially an exercise of causal inference helps to understand that measures of equality of opportunity, or relevant indicators of circumstances, are not necessarily transportable across outcomes and contexts. While a measure might be a just measure of effort for one outcome, it may not necessarily be so for others. Moreover, this observation may be even more relevant when it comes to weeding out possible confounders of estimates.

Education as an outcome

Much of the research on inequality of opportunity measures outcomes by income or occupation. (Atkinson, 2015; Smeeding, Jäntti & Eriksson, 2011) The literature is vast, and as our focus is not on equality of opportunity for income or welfare, we will not provide a general overview of the literature, except for what emerges as necessary for the questions on causality and definition of types.

In the field of education, most studies using education as an outcome use measures of formal educational attainment, such as years of schooling. More recently, with the increasing availability of large-scale survey data, measures of learning outcomes have been used to measure equality and equality of outcomes. (Meschi & Scervini, 2014; OECD, 2013b; Pfeffer, 2015) Most research has focused on individual data sets, quite often producing data on equality or equality of opportunity as a part of a broader set of primary analyses of the results of the survey. However, several studies have recently emerged that are focused on equality of opportunity, using learning outcomes data from a large number of national and international surveys. Reardon (2011) has used a wide variety of national and international data sets to show increasing achievement gaps in the United States from the 1960s to 2010s. Chmielewski (2019) has, in turn, shown similarly increasing gaps internationally in most countries, using a comprehensive set of 30 large-scale surveys conducted over a period of over 50 years since the early 1960s.

As will be argued below, the suitability of different specifications for measurement of equality of opportunity depends on the outcome to be measured. Thus, while the underlying schema for measuring inequality of opportunity may be the same irrespective of the outcome, the outcome affects where the responsibility parameter should be set.

Outcomes to expect

After the causal model is identified, the model can produce a counterfactual distribution by replacing individual values with the values predicted by the model. There are two main methods of valuing the opportunity sets by type. The dominant one takes the average outcomes of a type or a tranche, while the other takes a reference outcome. In approaches using reference outcomes/circumstances, the results obtained depend on the chosen reference. Usually the reference is chosen to, e.g., fulfil the compensation rule known as egalitarian-equivalence, which sets the reference circumstances so that individuals with reference value circumstances face only the reward principle and the redistribution from individuals with above-reference circumstances is equal to the redistribution to the individuals

with below-reference circumstances. A combination of the choice between average/reference outcomes and the above-presented choice between ex-ante and ex-post strategies perspectives is presented in ***Table 1.1***, which presents most of the commonly used strategies for measuring equality of opportunity.

The two main ex-ante measurement approaches are *Between-types Inequality* and *Direct Unfairness*. Both approaches share the interpretation that outcome differences between types are based on circumstances and approach the problem by smoothing out differences in outcomes that are due to effort, to arrive at a counterfactual distribution reflecting only circumstances. *Between-types Inequality* replaces individual outcomes with the average outcome of the type. (Bourguignon, Ferreira, & Menéndez, 2007; Checchi & Peragine, 2010; Ferreira & Gignoux, 2014; Peragine, 2002) The average level of an outcome can readily be thought of as the expected value of the outcome predicted in the absence of information on effort. A close relative of that approach is *Direct Unfairness*, which replaces individual outcomes with a reference outcome of the type. (Fleurbaey & Schokkaert, 2009) In both models, outcomes are predicted exclusively by unjust causes so that the distribution produced is considered unjust.

The counterfactual distribution amounts to smoothing the distribution, as all within-type variation, representing just variation of outcomes with effort, is removed. For example, the opportunity to attain tertiary education is measured by the probability of attaining tertiary education conditional on circumstances. The difference between equality of outcomes and equality of opportunity is then the difference between the distributions of outcomes and the distributions of expected (values of) outcomes.

The counterfactual distribution purified of individual differences in outcomes within types is thus purified of the effect of effort and should, if the identified causal model is correct, measure the contribution of circumstances to the distribution of outcomes. As the differences due to effort have been removed from the smoothed distribution, the remaining inequality of distribution measures inequality of circumstances and thus inequality of effort.

The two main ex-post measurement approaches are *Within-Tranches* and the *Fairness Gap*. In *Within-Tranches* (Checchi & Peragine 2010) the amount of effort is calculated by replacing each individual outcome in a given tranche (such as the 37^{th} effort percentile) with the ratio between that outcome and the average outcome in that tranche. By contrast, in the *Fairness Gap*, each individual outcome is replaced by a ratio of that outcome and the outcome achieved with the same level of effort in a defined reference circumstance. Inequality of opportunity is here constituted by differences in outcomes between individuals who are in the same tranche by sharing the same level of responsibility characteristics, i.e. effort.

In all of the approaches, observed conditional outcomes are used to infer counterfactual outcomes. Thus, the statistical expected value of outcome given type is, explicitly or implicitly, interpreted causally as outcome given treatment. (Lefranc, Pistolesi, & Trannoy, 2009; Roemer & Ünveren, 2017)

Table 1.1: approaches to constructing a counterfactual distribution

	average	reference
ex-ante (types)	Between-Types Inequality	Direct Unfairness
ex-post (tranches)	Within-Tranches Inequality	Fairness Gap

3.4 Measuring distributions and effects

Inequality of opportunity is measured by measures of the counterfactual distribution produced by the causal model. But the distinguishing feature of equality of opportunity is the use of a counterfactual distribution purified of individual differences in effort, leaving only the distribution of outcomes associated with, and supposed to be due to, background characteristics or circumstances beyond individual choice or control, i.e. inequality of opportunity.

The equality of the counterfactual distribution can be measured in various ways, just as the equality of any distribution can. However, the specific nature of equality of opportunity can be brought to bear on the choice of the measure of the distribution, as we shall presently argue.

Direct and indirect measures

Even when measuring a distribution and choosing from among various measures of inequality as dispersion, there is a choice to be made vis-à-vis measuring the distribution as is or, again, relative to something. Ramos and van de Gaer (2016) divide measures of equality of opportunity into direct, indirect and norm-based measures. Direct measures are those that simply measure the inequality of the counterfactual distribution. Indirect measures measure the unjust counterfactual distribution relative to the observed distribution of outcomes to determine how large a share of inequality of outcomes is due to inequality of opportunity. Norm-based measures, in turn, calculate for each individual a norm-based outcome and measure inequality of opportunity using the distance between the norm-based and observed outcomes.

The terminology of directness-indirectness is somewhat unfortunate in a context where opportunities are indirectly observed, as it can create confusion, but the criteria for the classification of measures are essential. While it is sometimes necessary to measure the simple extent of inequality of opportunity, we might sometimes be interested in the relative contribution of inequality of opportunity to overall inequality of outcomes. But we agree with Ramos and van de Gaer that the indirect measure cannot be interpreted as a norm-based approach because in the norm-based measurement it matters who gets what, while in the indirect method it does not; this makes the indirect method unattractive as a measurement of inequality of opportunity.

More importantly for our practical purposes, we agree with Ramos and van de Gaer (2016) that the indirect measures that measure unjust inequality relative to overall inequality do not capture what is normatively pertinent. As they note, 'for true opportunity egalitarians, those concerned with equality of opportunity rather than equality of outcome, the answer to the question is irrelevant.' What is important is the amount of unjust inequality, and that is not affected by the extent of just inequality. This has led us to choose a direct measure to measure equality of opportunity and reject indirect measures that measure the relative contribution of inequality of opportunity to the overall inequality of outcomes. We note, of course, that such an approach has been suggested, in the context of education, by Ferreira and Gignoux (2011, 2014), and used by Natkhov and Kozina (2012).

Ramos and van de Gaer (2016) rightly observe that the 'theoretical basis for many of the inequality measures that are used remains rather weak', as many of the technical measurement choices have often been made on a rather ad hoc basis, without grounding them in theory. This relative lack of grounding would justify expending significant effort to describe the substantive interpretation of

the measures used and the background assumptions that were accepted when the methods and measures were chosen.

Inequality as dispersion

After choosing to measure inequality directly, many measures of inequality as distribution are then available to measure inequality in the distribution of opportunities. In the empirical literature, equality of outcomes is sometimes referred to as 'inequality as dispersion' (Werfhorst and Mijs 2010), because the distribution of outcomes at any given moment may be measured using statistical measures of dispersion such as range, standard deviation, variance, coefficient of variation, Gini coefficient or Atkinson indices. (Atkinson, 1970) Different measures emphasise different aspects of the distribution, and thus the choice of the measure depends on the questions asked.

Temkin (1993) has identified 12 normative aspects of inequality, arising mostly from different combinations of three different principles of equality and three different views of complaints. Views of complaints refer to the point of comparison adopted for the considerations regarding inequality. The *average view of complaints* considers only individuals falling below the average level to have a complaint, while the *relative to the best-off person view of complaints* considers that falling short of the level of the best-off person is grounds for complaint in a comparison of equally deserving individuals. The *relative to all those better off view of complaints* accepts that all but the best-off have a complaint, but the size of the complaint depends on how the individual compares relative to all the others who are better off than she is. These views of complaints can then be combined with different principles of equality. The *maximin principle of equality* is the Rawlsian principle of maximising the average level of the worst-off group, the *additive principle of equality* considers inequality to consist of the sum of individual complaints, and the *weighted additive principle of equality* that in this summing of complaints larger complaints merit extra weight. These principles and points of view may be combined in different ways, to produce nine of the twelve aspects of inequality identified by Temkin.

Temkin (1993) interprets that, instead of showing equality to be hopelessly vague and contradictory, the multitude of principles and points of view shows that equality is multifaceted and that 'various measures are best regarded as measures not of inequality itself, but of certain *aspects* of that notion.' Different statistical measures capture these different facets of inequality to different degrees. For example, range matches the combination of *relative to the best-off person view of complaints* and the *maximin principle of equality* while the relative mean deviation combines the *average view of complaints* with the *additive principle of equality*.

A statistical measure needs to be chosen to capture the relevant aspects of inequality. The following presentation will focus on the achievement gap, used in the later empirical analyses. Some additional remarks will be offered on the socioeconomic gradient, which is also often used in empirical work, as well as on the possibility of measuring dispersion using some Gini-type measure.

Range - the achievement gap

Now, since the ex-ante approach 'views inequality of opportunity as reflected in the degree to which average outcomes of different types differ' (Roemer & Trannoy, 2016), we may want to seek a measure that compares the average attainment of different groups with each other. Here, one traditional and well-established measure presents itself: range.

The achievement gap between advantaged and disadvantaged groups – i.e. range of group averages - is one of the most traditional measures of educational opportunity and has been widely used since Coleman (1966) showed the size of the racial achievement gap in the USA. The simple nature of the metric has made it widely used, and the Handbook on Measuring Equity in Education considers that

> in many cases, the easiest and most accessible analysis of impartiality involves simply presenting statistics disaggregated by different groups in a table or graph. Tabulating the gaps or differences between particular groups, such as the difference between the richest and poorest, enables comparisons to be made across countries or over time. (Omoeva, Muoussa, & Hatch, 2018)

The extensive use of the achievement gap is partially due to the low data requirements that it has, requiring little data beyond average levels of outcomes by group.

Ever since the Coleman Report, and even before it, these characteristics made the achievement gap a very commonly presented measure of equality of opportunity, even though it is often not presented as a measure of equality of opportunity – a fact that may have contributed to many a misinterpretation of measures of opportunity as ones of outcomes.

The measure is standardly used in all large-scale surveys, national or international, either directly or in the related form of the socioeconomic gradient. Socioeconomic gradient measures the impact of background by the slope of the regression line of outcomes regressed on background characteristics. It approaches the achievement gap as the number of indicators and their possible values to measure background decrease. The two measures become identical when the background is measured using only one indicator with two values. The improvement of outcomes with an improvement of background by one class thus represents the difference between the high and low types. For both measures, a low number of indicators contributing to type lowers data requirements.

OECD research using PISA or PIAAC data does not refer directly to equality of opportunity, but in effect the OECD uses both univariate measures of equality of outcomes, such as standard deviation, variance or the difference between 75^{th} and 25^{th} percentiles, as well as bivariate measures of equality of opportunities, such as the slope and the coefficient of determination of the socioeconomic gradient. (OECD, 2007, 2010, 2013a, 2013b)

In the context of counterfactual inequality of opportunity, technically the achievement gap produces a counterfactual distribution using the between-types inequality approach and then measures the inequality by range, the distance between the best-off and worst-off individuals in the distribution. Because individuals are ranked by the expected values of outcomes, the best-off and worst-off individuals are stand-ins for the average outcomes of the groups with best and worst circumstances, respectively.

The use of the achievement gap as a measure of inequality of opportunity can also be justified by the wide acceptance of *relative to the best-off person view of complaints* and the *maximin principle of equality*, as it is the only statistical measure of dispersion that combines the two. (Temkin, 1993)

In essence, it embodies the principle that not only those below medium or some other such level, but all who fail to get the best that can be provided, have grounds for complaint, and combines it with the principle that justice should maximise the outcomes of the worst-off. As we are concerned with differences in outcomes that can be explained by circumstances rather than effort, the entire difference in outcomes is undeserved and can be considered unjust.

This Temkinian approach has been visible in the ex-ante approach, associated with van de Gaer (1993), in which the opportunity set faced by a group has in practice been predominantly measured using the mean income (or achievement) of the type. Since equality of opportunity requires the equal value of opportunity set in all circumstances, differences in means between types are a measure of between-type inequality. Or, as Roemer (2006b) expresses the same reasoning:

> Now, to be more precise, we are not simply concerned with equalizing distribution functions but rather with equalizing them at the highest possible level. To avoid a discourse on what it means to equalize a set of distribution functions, let us simply summarize such a function by its mean. This leads immediately to the concept of maximin: that is, an opportunity egalitarian should seek to maximize the mean advantage level of that type with the lowest such mean.

Between-type range as the effect size of the treatment

The primary justification for using range as the appropriate measure of inequality of opportunity in the following analyses draws on the causal interpretation of the results. As the difference between types is interpreted as the difference between treatments, the range of outcomes of observed groups becomes interpreted as the difference between counterfactual outcomes, or more precisely: between two treatments. Here we draw from causal calculus, or the mathematically corresponding potential outcomes framework or counterfactual approach sometimes referred to as the Rubin Causal Model. While the approach is mostly used in epidemiology and increasingly beyond it, it began in the setting of educational testing, as a solution to the challenges of causal inference with observational data. (Rubin, 1974, 2005) In the counterfactual approach, the main measure of the effect size of a treatment is the average treatment effect or the average causal effect. The average causal effect is measured by the difference in the mean outcomes of the treatment and the non-treatment groups and it is the most commonly used measure of population causal effect. (Hernán & Robins, 2020)

Thus, the achievement gap is a measure of the effect that would be achieved if a disadvantaged group were to receive the same treatment as the advantaged group. In the case of the empirical analyses in the latter part of this work, it is the difference between the treatment of individuals who have parents with high education instead of low education. As a reminder, here it makes a difference that we want an estimate of an ill-defined treatment rather than a well-defined, actionable handle that is closer to understanding the mechanism of the causal link. It is by definition true that, if the observed effect is one of a hypothetical treatment that we do not know, but that would make the untreated group like the treated one in all causally relevant respects, then the average treatment effect is the observed mean difference between the types. But, as noted above, '[t]he implicit intervention […] is likely to be too difficult to implement to be meaningful to policy makers.' (Hernán & Taubman, 2008) This is especially the case if a significant part of the treatment turns out to be genetic.

The interpretation of the achievement gap as effect size is rarely underlined in the literature, which more often makes reference to, e.g., the accessibility of the metric. (Omoeva, Muoussa, &

Hatch, 2018) This may be due to unwillingness to engage with causal language when measuring inequality of opportunity. But as inequality of opportunity is inequality of causes of outcomes, the questions of causality cannot be avoided. It is the effect size of the treatment represented by the type if the causal model is identified correctly. But the distribution measured is not that of opportunity if the causal model is incorrectly identified. Thus, avoiding causal language in the presentation of the results merely adds to murkiness and confusion. Here, causal language is used, but the uncertainties involved are also clearly spelt out. The measure gives an intuitive interpretation of the evolution of a system as it moves along the ability-opportunity curve (to be introduced in ***Section III***) or when the shape of the curve changes.

This metric has been used for international comparison of the effect of parental education on learning outcomes (Goodman, Sands, & Coley, 2015), and metrics of the same construction have also been used, with PIAAC data, to measure differences in income, university access and years of schooling associated with differences in parental education. (Jerrim & Macmillan, 2015) The same measure has been previously used with PIAAC and IALS data, by Green et al. (2015), who refer to the measure as 'social origins gradient', but measure social origins using solely the three-level classification of parental education and define the gradient as the 'the point difference in scores that can be predicted for an individual when the education level of his or her parent(s) is increased from the bottom unit to the top unit.' This statistical prediction of expected value given type is thus given a causal interpretation as it is used as a measure of 'how much family background influences the level of an individual's skills'.

Population measures

This is where our choice of measure differs from measures that would be more interested in the evolution of opportunities in the entire population. Between-types achievement gap measures the magnitude of the difference in treatment between best- and worst-off groups. With the focus on extremes of the distribution, the achievement gap loses much information on the shape of the distribution between the extremes. Indeed, it is entirely insensitive to any changes in the middle of the distribution. If we wanted to measure aggregate inequality of opportunity in the whole population, we could measure, for example, the Gini coefficient of counterfactual learning outcomes.

Such a measure would suffer from considerable limitations because learning outcomes are not cardinal measures, and the scale of their measurement has no true zero. But the measure would be affected by changes in the middle of the distribution and improving or declining relative position of the middle parental education type would affect the level of measured inequality of opportunity. Moreover, the size of the high and low parental education groups would affect the result, and inequality of the distribution would be reduced by a decrease in the size of the low parental education type.

What would be lost with such a measurement, in turn, would be the focus on the difference of the treatment effects of the relatively best- and worst-treated groups in society. What is missing is an understanding of how well off the well-off are compared to the badly-off. In our empirical approach the focus is on the treatment effects of social background, and thus the achievement gap is the natural choice of measure. Indeed, our use of range as the measure of inequality of opportunity is dictated by our wish to engage with the question on the relationship of innate ability and social circumstances in

affecting measured inequality of opportunity. This would not be achieved with e.g. Gini of counterfactual learning outcomes, because the effects of the changes in the middle of the distribution would affect the results and thus break the clear link between the groups that range offers as a measure of inequality.

Summary - achievement gap as measure of equality of opportunity

Measurement of equality is an exercise in ill-defined causal inference, where the variation of outcomes *with* background circumstances is taken to imply that outcomes vary *due* to the background circumstances. When we take the differences between types to be ill-defined interventions, the difference between two types is conceived of as a hypothetical treatment that would turn one type into the other in all causally relevant respects. When treatment is ill-defined, variation with type becomes variation due to type, by definition.

Background characteristics may be measured on an absolute or a positional scale, as most indicators of circumstances have both absolute and positional characteristics. Sociological research has traditionally measured educational attainment in absolute terms but has recently turned more towards positional measurement. While educational attainment is also positional, education does have absolute effects on learning outcomes and productivity. We follow sociological tradition and primarily measure parental education on an absolute scale but supplement the absolute measurement with a positional measurement.

Inequality of opportunity will be measured by the achievement gap between high and low parental education groups, as measurement of equality of opportunity takes type to be a treatment and the primary measure of effect size is the difference in mean outcomes between the treatment and the non-treatment groups.

Section II – Innate ability and responsibility

4 Norms, principles of justice and rules of regulation

Normative assessment is the application of norms or normative facts about how things should be to how they actually are. Any normative assessment of any issue is based on two types of arguments: normative and natural. The latter are assessments of the natural state of affairs that are made relevant by the normative facts on which the assessment is based. Thus, ultimately, the definition of types based on circumstances is a normative question.

The standard methods for estimating inequality of opportunity usually explicitly endorse some version of Temkinian canonical egalitarianism, which finds inequalities unjust if they arise through no choice or fault of the individuals concerned. This choice is neither innocuous nor straightforward, however, because it leaves the question of innate ability untouched. Innate ability is quite obviously beyond the control of an individual, and thus morally arbitrary. Most empirical approaches take individuals to differ in their innate ability and those abilities to not-unjustly influence individual outcomes. Instead of explicitly dealing with the normative and empirical implications of that choice, the Gordian knot is cut by noting that, in light of the measures used, individuals are responsible for their innate ability when it is independent of their background and not responsible for it when it is associated with their background.

Because the question of innate ability is central to any assessment of inequality of learning opportunity and to any policy question related to the relationship between equality of opportunity and the level of outcomes, it is necessary to clarify different types of responsibility characteristics. Before we turn to normatively assessing different causes identified by causal inference, it is necessary to take a brief look at the fundamentals of normativity. The presentation is by no means extensive but draws attention to some critical questions of normativity that are pertinent to the question at hand. First, we make the distinction between normative and natural facts, which is used to understand the difference between fact-sensitive and fact-insensitive normative principles and to clarify the difference between fundamental principles of justice and rules of regulation.

Causal and moral responsibility

While the moral responsibility of individuals for the causal consequences of their actions is an essential justification of moral responsibility, causal and moral responsibility are conceptually independent. In causal responsibility, individuals are *deemed responsible* for something and in moral responsibility, they are *held responsible* for something.

Causal responsibility is fundamentally an empirical question. There are complications involved when it comes to observing causality, but if there are causal processes in the world, then they can, at least in theory, be identified with a sufficient amount of the right type of data. It may not be possible in practice to collect such data, but we may at least identify the criteria that such data would need to fulfil.

Moral responsibility is complicated in a different way. It is a question of what individuals are held responsible for as a matter of justice. All conceptions of justice agree on holding individuals responsible for some but not all of their outcomes. While extremes are logically possible, holding individuals responsible for whatever might happen to them is not a position that finds much support,

any more than the position that all outcomes should be fully equalised. If either position exists in the literature, it exists as something argued against, rather than something defended. As we saw, both the critics of egalitarianism – who endorse responsibility as a necessary consequence of freedom – and the responsibility sensitive egalitarians – who want to hold individuals responsible for their choices, preferences or control – link moral responsibility foremost to causal responsibility. But they differ fundamentally on how the *responsibility parameter*, the dividing line between equality and responsibility, is set.

Responsibility parameter

The crucial choice for any theory of equal opportunity is the setting of 'the degree of responsibility of persons as a parameter in a theory of equality.' As Roemer and Trannoy (2015) note,

> Once one assigns a value to this parameter, then one has a particular theory of equality of opportunity, because one then knows for what to hold persons responsible.

This parameter may be set in multiple ways, from holding individuals responsible for almost everything to holding them responsible for almost nothing; the concept covers the whole range of different conceptions of justice. The responsibility parameter may be treated as an exogenous variable, as a parameter that, for Roemer and Trannoy (2015), is

> supplied by each society, which has a concept of what its citizens should be held responsible for; hence there is a specific theory of equality of opportunity for each society, that is, a theory that will deliver policy recommendations consonant with the theory of responsibility that that society endorses.

This approach is political and practical rather than philosophical in the sense that it takes any conception of responsibility characteristics as exogenous, given by values of society, without inquiring into the normative justification of that conception. It is indifferent to whether the normative system is based on irreducible normative truths or ultimately an illusion. If we take a Parfitian view that there are normative truths, there can be a single, just responsibility parameter that we simply must find. If we take normativity to be fundamentally an illusion, we can choose the responsibility parameter at our whim. Whatever the basis for the parameter, no theory of justice or conception of equal opportunity can do without it.

Interpretive concepts and unity of value

The idea of an interpretive concept is especially useful for the treatment of a conception like equality of opportunity that seeks to combine equality with individual responsibility and freedom. After all, the supposed trade-off between equality and freedom is very often discussed, as equality is seen to require direct or indirect equalisation that effectively limits the liberty of individual action. (Berlin, 1958; Nozick, 2013)

Now, it is clear that the negative liberty of an individual will have to be interfered with if equality, of almost any kind, is sought between individuals. If liberty is taken to be freedom from any interference, then the state abridges liberty all the time. But we are interested in freedom as a normative concept, so the question is, do we 'compromise… an important value when we do so?' Freedom,

in a neutral sense, is limited when arson and rape are prevented. But does that limit the scope of freedom as value? Is something of value lost when the liberty to commit rape or arson is limited? The Dworkinian answer is that when we refer to liberty as a value, that value does not include the liberty to commit rape and arson. Liberty considered as a value is not infringed by preventing rape and arson, and no valuable liberty is lost if rape and arson are forbidden by law. (Dworkin, 2011)

When we treat central value concepts as interpretive concepts, we quickly find a unity of value in which values are not in conflict but require mutual adjustment. The unity of value approach contends that normative questions require normative arguments and a conflict between two normative values can only be resolved by a normative argument. And if the conflict can be settled by a normative argument, then the two normative values cannot be in conflict within a single normative system of value. (Dworkin, 2011)

In this conception, equality and liberty are interpretive concepts and, rather than finding a balance between the two, deciding how much of each to sacrifice to achieve some of the other, the task is to find definitions of both that are mutually compatible and capture what we consider to be valuable in liberty and equality. As only normative arguments are valid in seeking a balance between different normative conceptions and building mutually consistent conceptions, seeking a balance between different conceptions always happens within a system of normative reasoning. Any apparent conflict between normative principles will either reveal the structure to be internally inconsistent, because it cannot be resolved with that normative system, or the conflict will be resolved by redefining the relevant conceptions in a mutually consistent way. In the end, the interconnected set of values at the core of a normative system must be free of internal contradiction. That is why a proper conception of equality of opportunity requires mutually consistent conceptions of equality, freedom and responsibility.

The normative is a realm of interpretive concepts because, unlike in the natural world, things cannot be barely true. In the natural, we can change the bare facts of the world, or imagine them to be different, and the changes do not ripple throughout the universe as in the normative. One could imagine a world in which the blue pen now on the table was red, and that would be the only difference between the two worlds. The colour of the pen is a bare fact. These facts are what observations of the world provide evidence of. With a sufficient number of observations, we can reasonably take specific natural facts as being true, even if we should remember that empirically nothing is ever entirely and undoubtedly true or untrue, but only provisionally so, and consequently to different degrees.

There are no bare facts in morals, or in the normative realm in general. It is impossible to think of a world that would be like this one in all other respects, but in which torturing babies for fun would not only be permitted but good. Norms form a structure that is interlinked, and a normative stance on the justification of torturing babies would ripple to other principles in the same normative system.

The unity of value approach may be understood, in rough analogy, as seeking a normative value equivalent of the social welfare function in economics, a type of general moral value function. Whilst the conflict of values approach interprets that freedom and equality are in conflict because neither can be sacrificed without sacrificing the other, the unity approach seeks consistency by treating equality or freedom as values to the degree that they contribute to maximising the value of the moral value function. If we do not lose something of moral value should we lose the right to commit rape and arson, this indicates that the freedom vector in the moral goodness function is not a vector of non-interference, where maximising non-interference would maximise freedom as a value. If we do not lose something of value when not everyone has the very same outcomes, then the vector of equality

as value does not maximise equal distribution of every outcome. The unity of value approach posits that once all values have been defined so as to be mutually compatible, individual value vectors, like equality and freedom, are no longer in trade-off and the value function is maximised when all the vectors contributing to the function have been maximised.

Justice, regulation and sensitivity to facts

Norms are built on natural facts and more fundamental norms. With a norm, it is always appropriate to ask the question: why? And the answer to that question consists of two parts: a more fundamental normative principle and a natural fact, which together justify the more specific norm. The more specific the norm, the more levels this structure has. As long as the question 'why?' can be answered with a combination of a norm and a natural fact, the normative part of that answer is a more fundamental principle.

This quest for an answer would end with either a fact-insensitive normative principle, which is irreducibly normative, or with an appeal to only natural facts, in which case one could get an ought from an is, and the whole structure would come tumbling down. Alternately, the case for fact-insensitive principles can be made without reference to the Humean principle of not deriving an ought from an is, and merely positing like Cohen (2008) that

> a principle can respond to (that is, be grounded in) a fact only because it is also a response to a more ultimate principle that is not a response to a fact. Accordingly, if principles respond to facts, then the principles at the summit of our conviction are grounded in no facts whatsoever.

Full and extensive treatment of the matter would clearly be out of place here, but an inescapable conclusion of this would seem to be that

> '[n]ormativity is either an illusion, or involves irreducibly normative truths.' (Parfit, 2011b)

Normativity can obviously be an illusion in very many ways, and Parfit's work (2011a, 2011b, 2017) is a thorough analytical presentation and criticism of all oft-espoused philosophical views, culminating in a defence of non-natural cognitivism, which takes normative claims as intending to state truths and that there are objective normative truths that are irreducibly normative. This conception of irreducible normative principles is in tension with the positions of, e.g., John Rawls (2009), Christine Korsgaard (1992, 1996) and Bernard Williams, which root normativity in human nature or the human situation.

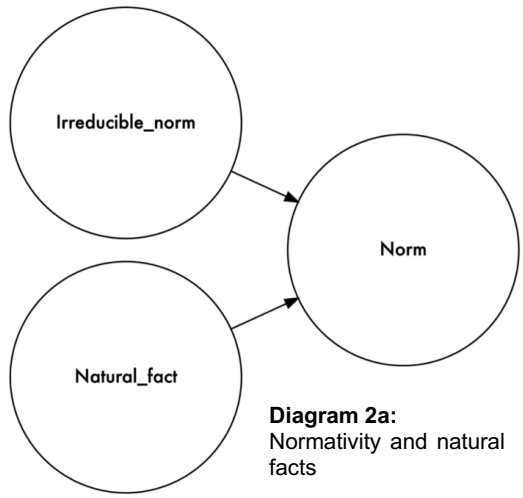


Diagram 2a: Normativity and natural facts

Here it is not necessary to address the question of whether these fundamental principles or irreducibly normative truths are truths or not, as it is enough to note that they are irreducibly normative, in the sense of not responding to any

non-normative facts. These irreducibly normative principles are a logical necessity in normativity as an independent domain of reasoning. We can believe that these irreducible norms are normative truths as Parfit (2011a, 2011b, 2017) and Dworkin (2011) believe and as this author is provisionally inclined to think. Or we can believe that normativity is fundamentally an illusion. In this view, there is no truth to be had on normativity. In this view, the world is a world of natural facts and normativity is not to be found from the world but to be imposed on it. In such a world, normative statements are ultimately a question of taste and preference. However, even in such a system, normativity derives power from not being wholly arbitrary and merely a matter of taste. Thus, we may find that different sets of normative statements may form several parallel normative systems that are internally consistent. If normativity is an illusion, ultimately there is no normative argument that can persuasively be made to choose between the different systems.

But, for all normative systems, there are irreducible norms, normative statements that are not derived from more ultimate norms, and thus represent what the normative system takes to be a normative truth. These, in essence, play the role of the *Grundnorm* in each normative system. The emphasis here is on the role of the basic norm as the highest norm to which a normative system may appeal. Like an irreducible moral truth, it may be not be created, as some deeper normative principle would have to give some entity the right to set norms. It may not be derived, as derivation would require something more fundamental. And it may not be questioned, as no appeal can be made to anything more fundamental to justify the basic norm. The basic norm and the irreducibly normative truth are the steps from which we take the leap from the world of what is to the world of what ought to be. (Kelsen, 1967, 1945) Even if we believe that several contradictory normative moral systems may exist with no way of establishing their relative merits, each of these systems would need to be based on some irreducibly normative principles that are taken as true by the system.[13]

Whatever the position we adopt, any normative system must, to be normative, contain normative statements that are taken to be true by the system and not derivable from anything more fundamental or from non-normative facts.

Principles of justice and rules of regulation

The difference between fact-sensitive and fact-insensitive normative principles allows us to clarify the distinction between two levels of principles of justice.

Rawlsian principles of justice are identified with the aid of a veil of ignorance, which makes agreeing individuals ignorant of their own position, including their external and internal endowment, in society, so that their judgement on rules to be followed would not reflect their own interest. As the emerging rules cannot reflect partiality or individual self-interest, since individuals are unaware of their particular interest, they are considered principles of justice by Rawls (2009).

[13] However, we should reject the position that we can be sure that universal normative truths do not exist because there is disagreement in normative questions. Parfit's treatment of the subject is convincing, but more importantly the standard objections to the impossibility of normative truths seem to be extremely weak. While it is clear that we do not know the ultimate normative truth any more than we know the ultimate truth of the natural world, it would be misguided to argue that in either area this fact, or the fact that different societies have entertained very different conceptions of how nature works or what is right or wrong, provides a sufficient reason to conclude that the truth is not out there.

The Rawlsian veil of ignorance has been criticised on many grounds as being too thick or too thin for several purposes, but here our interest is in Cohen (2008), who argues that the veil of ignorance is too thin to allow individuals to identify principles of justice.

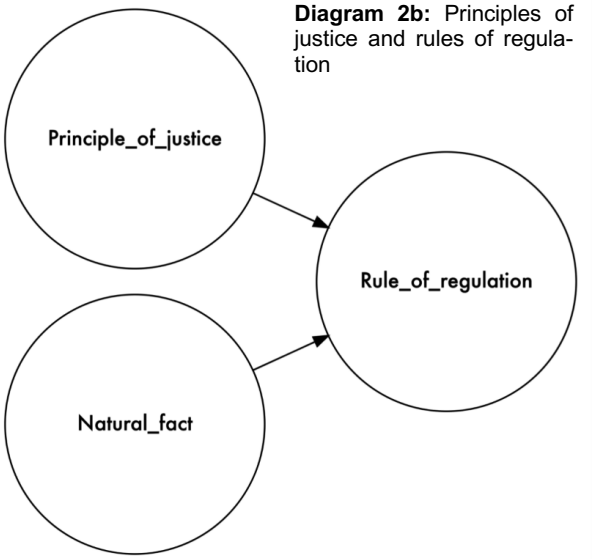


Diagram 2b: Principles of justice and rules of regulation

> For Rawlsian constructivism, fundamental principles of justice, for all that they are fundamental, which is to say, not derived from still more fundamental principles, reflect facts. Rawls believes that because he misidentifies the question 'What is justice?' with the question 'What principles should we adopt to regulate our affairs?' For facts undoubtedly help to decide what rules of regulation should be adopted, that is, legislated and implemented, if only because facts constrain possibilities of implementation and determine defensible trade-offs (at the level of implementation) among competing principles.

Cohen's terminology allows us to differentiate between principles of justice and what Cohen refers to as *rules of regulation*, the rules to be implemented as rules of social interaction. *Rules of regulation* are *principles of justice* modified by any feasibility considerations that might apply in a given situation. They are not principles of justice, but fact-sensitive normative principles, amounting to realisable principles of justice under a set of practical constraints.

Rules of regulation evidently rely on natural facts in addition to normative principles. In Cohenese, rules of regulation are fact-sensitive normative principles, while principles of justice are fact-insensitive normative principles. Achieving justice in the sense of fact-insensitive fundamental principles of justice might be unfeasible for non-normative reasons. As far as these represent natural constraints on how social systems are organised, they may legitimately be reflected in the shape of the rules of regulation to be adopted, but that does not alter what justice is. The fact that a principle of justice might be unrealisable does not change what would be just, it simply makes the just unattainable. Temkin (2011) refers to such universal principles of justice when he discusses natural injustice. He notes that it is naturally unjust if 'despite the best possible personal and social responses to her situation' by a saint, 'her life is one of unremitting pain and misery'.

The Rawlsian principles of justice are rules of regulation, as the choosers of the rules beyond the veil of ignorance have information on the society for which they are seeking principles, even if they are ignorant of their endowments. Thus, they can take into account practical limitations that implementation of fact-insensitive principles of justice might face. For example, we may have moral principles that lead us to think that the world would be more just if all outcomes were within reach of all individuals, depending only on their actions. But we know that humans possess biological variation in their innate characteristics, which affect the distribution of possible outcomes, at least in the near to medium future. Even if principles of justice might not accept outcome differences based on

innate differences as legitimate, rules of regulation can hardly choose not to take into account variation in innate ability and the practical difficulty of redistributing natural endowments.

In this conception, principles of justice are universal. Rules of regulation are affected by the objective material constraints present in the world in which the rules are to be applied. These material constraints range from regional, relating, e.g., to the availability of different crops or minerals, to the universal, relating to the universal laws of nature. On all levels, the objective bare natural facts pose material constraints on the realisation of principles of justice and the just rules of regulation rise from a combination of facts and fact-insensitive normative principles. As with all things factual, we may always ask counterfactual questions and thereby find more ultimate normative principles that would apply if the brute natural facts of the world did not limit the range of their application. Under such reasoning, the universal moral truths may be universal beyond this universe of ours and extend to worlds that might obey completely different natural laws.

But why should we concern ourselves with such hypothetical universal justice, even if we believe that fundamental moral truths exist? Temkin (2011) gets the gist:

> What is the point of calling such a situation unjust? It tells us something important about the situation. It tells us that if, contrary to fact, we *could* do something about it, then, *ceteris paribus*, we *should* do something about it.

Social and natural freedom

The difference between the *neutral* and *rights accounts of freedom* (See subsection *Freedoms* on page 31) cuts deeper than only freedom. The neutral account of freedom takes as limitations on freedom any constraints or limitations that are placed on our actions by the actions of other morally capable agents. The rights account of freedom brings in the normative system in force and takes it that when the law gives the state or individuals rights that limit the actions of individuals, those limitations are not limitations on freedom. We may take it that the difference between the two is that the neutral account refers to *descriptive freedom of action*, while the rights account refers to *just freedom of action*. Just as all descriptive inequalities are not normatively unjust, not all forms of descriptive freedom are normatively unjust. What both conceptions share is the view that freedom is a social concept that can only be limited by the actions of morally competent agents. They merely differ in their view on whether just interference by a morally capable agent amounts to a limitation of freedom (neutral account) or is part of the definition of freedom (rights account).

There is, however, another relevant dimension to freedom, which is the distinction between social freedom, limited by the actions of other morally competent agents, and natural freedom, limited by the laws of nature and our technology. It may be noted that, under these terms, the neutral account of freedom corresponds to natural freedom at any given level of technology. It does not take any limits on our actions that are posed by the laws of nature and by our current level of technology as

Table 1.2: Natural and social freedom

	Universal	Limited
Social freedom	Neutral account of freedom	Rights account of freedom
Natural freedom	Omnipotency	Neutral account of freedom

limitations on freedom but takes all cases when moral agents limit the actions of others as limitations on freedom. In the scale of natural freedom, universal, unlimited freedom corresponds to omnipotence, where there are no natural limitations at all on the actions of an individual.

Why is this distinction useful? After all, we are very far from omnipotence, and the natural limitations on our actions are very much here to stay. The usefulness of the distinction comes from noticing that in the state of omnipotence, only fact-insensitive principles of justice are relevant because in the state of omnipotence there are no facts that principles would need to be sensitive to. There is nothing to be accepted as a given and to be taken into account when designing rules of regulation.

Distance from omnipotence

We may note that because of our distance from omnipotence – as individuals, societies and species – changes, the content of appropriate rules of regulations changes while fact-insensitive principles of justice do not change. The things that we simply must accept as irredeemable natural injustices change as our ability to manipulate the material world around us changes. Thus, the content of the rules of regulation is historically contingent, as it is, for us, contingent on the need to act as rules of regulation for a medium-sized social mammal on the outskirts of a smallish galaxy.

Making the distinction between rules of regulation and principles of justice may help us in the context of learning outcomes because it makes it easier to recognise the boundaries on which decisions are made on holding individuals responsible. The need to hold individuals de jure responsible for outcomes through rules of regulation, because we are de facto not able to help them with their predicament, may be easier if we recognise that we do not consider the consequences of their predicament as just (as principles of justice), but as currently irredeemable natural injustice that is to be considered unjust and redeemed when it becomes redeemable.

Equalities of access and of opportunity

The difference between fact-insensitive principles of justice and fact-sensitive rules of regulation also allows us to clarify the terminology of opportunity. From an individual, or principled, point of view, we may think that innate ability contributes to the opportunities available to us in a technical sense. But, as Cohen (2011) noted,

> We would not normally regard meagre personal capacity as detracting from opportunity. Your opportunities are the same whether you are strong and clever or weak and stupid: if you are weak and stupid, you may not use them well-but that implies that you have them.

Here, again, we can follow Cohen's terminology by adopting a distinction between access and opportunity. They describe the same thing, the possibility for an individual to pursue and achieve specific outcomes, but they do so in the contexts of principles of justice and rules of regulation. We may think of equality of access as a principle of justice that includes innate ability, while opportunity is a term of rules of regulation.

This distinction is not usually made in the literature, and it can be argued that making the distinction would have contributed to the presentation and clarification of some of the positions adopted

in the literature. We may take an example from the Handbook on Measuring Equity in Education that refers to the standard conception of inequality of opportunity because

> it is not clear that the philosophical concept of equality of opportunity is the right term for this concept in education. Equality of opportunity implies holding people responsible for the things within their control and not for circumstances beyond their control; but the moral basis for holding children responsible for their own innate talent or motivation appears weak. For this reason, we refer to this type of equity concept in education as impartiality. (Cameron et al., 2018)

This is unfortunate, because as far as differences in outcomes are due to innate ability, which is a non-alienable resource that cannot be redistributed, any equalisation of outcomes may only be achieved through levelling down. Most conceptions of justice would arguably find it unacceptable to pursue justice by levelling cognitive outcomes, so it might have been better to use the distinction between opportunity and access to state that

> Equality of opportunity implies holding people responsible for the things within their control and not for circumstances beyond their control. The moral basis for holding children responsible for their own innate talent or motivation appears weak, and is only justified to the degree, and by the fact, that they beyond our influence. This is how equality of opportunity differs from equality of access, which also requires that individuals are not held responsible for causes beyond their control. (modified from Cameron et al., 2018 quoted above)

Truly universal Kantian values

It is not necessary here to go into greater detail as to how rules of regulation should be set. However, it may be useful to note that while the Rawlsian veil of ignorance is too thin to help identify fundamental principles of justice, it may be in some respects too thick for choosing the right rules of regulation. Nozick (2013) takes the veil to be so thick that it biases the setting to produce principles of distribution that distribute goods as manna from heaven. Roemer and Trannoy (2015), on the other hand, find the veil too thick to choose just rules of regulation because, while Rawlsian agents behind the veil were 'assumed to know the laws of economics, and to be perfect agents of their self-interested principals', they were unaware of the life plans of their principals or the distribution of either the life plans or of 'physical and biological endowments in society'. These distributions are thus, justifiably, taken by Roemer and Trannoy to be the types of natural facts that affect the just rules of regulation.

One may argue that fundamental and fact-sensitive principles of justice are truly universal in the Kantian sense that arguably surpasses the current universe and are valid irrespective of the laws of nature and of any other natural characteristic of the world. Rules of regulation are then the practical applications of those universal principles of justice to take into account the natural facts of the laws of nature of the current universe[14], any pertinent natural and material facts of the more immediate surroundings of the moral agents in question, and the limits that their technological stage of development imposes on their social arrangements.

What Rawlsian or Dworkinian hypothetical agreements and insurance schemes do is to try to arrive at principles of justice by removing, by means of a veil of ignorance, information on the world

[14] Or a part of universe, if the laws are not universal.

and the individual that the negotiating souls represent behind the veil of ignorance. The thickness of the veil has implications for what questions the setting can feasibly answer. For some purposes, the Rawlsian veil may be too thick, but for arriving at principles of justice, it is too thin. Thus, supposedly, a thick enough veil of ignorance might be used to find truly universal, fact-insensitive principles of justice, that would not be affected even by the laws of nature.

5 Innate ability and responsibility without freedom

From the perspective of justice, we may think that god- or universe-given innate ability is one aspect of opportunity, as a lack of cognitive potential most certainly translates into a limited opportunity set. The literature is somewhat ambivalent on this point. Philosophical and sociological treatments usually agree that individuals should not be held responsible for causes beyond their control. In practice, innate ability is, however, more or less accepted as a responsibility characteristic, and no one seems to call for equalisation of outcomes due to variation in innate ability. Then again, standard treatments of equality of opportunity do not explicitly treat all innate ability as a responsibility characteristic, but in practice hold individuals responsible for at least some of their innate ability.

Before proceeding to how we should understand innate ability in relation to equality of learning opportunity in the next two chapters, this chapter sets the scene by taking a quick glance at treatment of innate ability in discussions on responsibility sensitive egalitarianism. Here, we attempt to justify in normative terms the choice to consider innate ability a responsibility characteristic and explore the relationship between normative constructs of equality of opportunity and the standard empirical measures of it.

5.1 Innate ability and merit

What if ability is not opportunity

The critical assumption in standard methods is the mutual independence of circumstances and responsibility characteristics. As noted, it is easy to defend when individuals are held responsible for causes within their control, as then all variation between types is legitimately due to either direct or indirect – through incentive effects on effort – effects of the circumstances.

However, when the choice is made to consider innate ability as a responsibility characteristic, the situation changes. Innate ability is not responsive to incentives and may itself cause parental socioeconomic status, education or income. When the innate ability is considered a responsibility characteristic, the background may not be automatically taken to be a treatment, as it is vulnerable to a systematic association of initial circumstances and innate ability.

This problem is illustrated with Diagrams 1c and 1d. They reproduce Diagram 1a but split responsibility characteristics in two, acknowledging innate ability as a possible cause of outcomes. Just like social circumstances, innate ability affects individual effort and outcomes. In addition, the modified diagram introduces the possibility that innate ability might have a causal role in determining the type. Thus, as a new assumption, the socioeconomic status, education or income of the parents may be due not only to social circumstances but to differences in innate ability. In treatment terms it indicates that part of the treatment of parental education or other SES measure is genetic and that the treatment differences between types may not be wholly eliminated without genetic treatment. As above, it may be of use to note that type is an analytic category. The genes of the child do not cause type when type is measured by parental education, income or some other circumstance characteristic, but type may be defined by genetic endowment. Here it is useful to remember that the causal model

describes a population, not an individual. The same model may be presented for individuals, distinguishing between parental and offspring innate ability, but that would add unnecessary complexity to the visualisation.

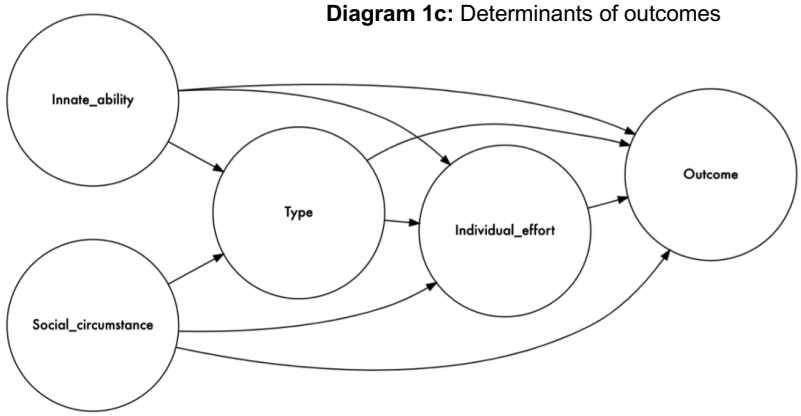


Diagram 1c: Determinants of outcomes

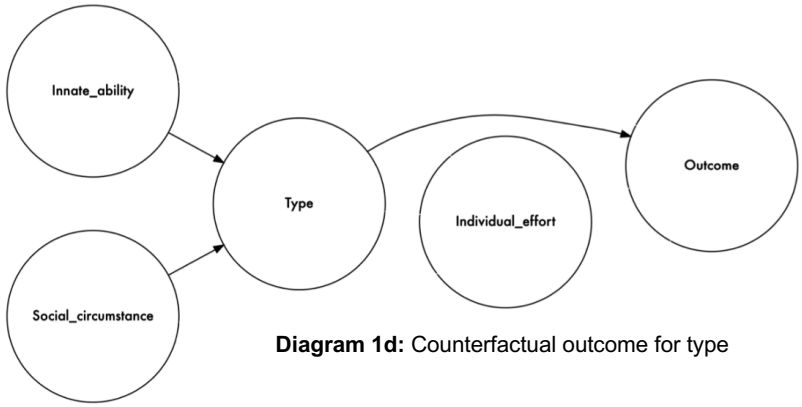


Diagram 1d: Counterfactual outcome for type

It is clear that when type by socioeconomic status also acts as a mediator on the causal path between innate ability and outcome, removing the direct causal path from innate ability to outcome (from Diagram 1c to 1d) is not sufficient to produce a measure of outcomes given circumstances, but only one of the outcomes given circumstances and ability associated with those circumstances. Differential innate ability thus acts as a confounder of socioeconomic status as a measure of circumstances. The problem for the empirical approach is that innate ability is as hypothetical and empirically elusive as social opportunities.

Given that innate ability is unobservable, the most straightforward strategy is to take the empirical measures as they are and simply state what the standard decomposition effectively holds individuals responsible for. The standard decomposition holds individuals responsible for innate ability that is *not* included in their circumstances and does *not* hold them responsible for innate ability included in circumstances. This cut between the two can be simply stated as a characteristic of the method of measurement and left at that. (Checchi & Peragine, 2010)

This is what all current empirical measures of equality of opportunity de facto do. Empirically, this is most likely because there is no alternative, as innate ability is unmeasurable. Partly it may also be because the normative underpinnings of the chosen empirical measures are often left unexamined in empirically-oriented work. (Roemer, 2012)

Moral arbitrariness of innate ability

First of all, there is widespread agreement that innate ability is morally arbitrary from the individual point of view because it is a product of a natural lottery, and does not give rise to individual merit or desert. This agreement seems to span the whole spectrum of philosophical views. The matter becomes more clouded when the philosophical discussion becomes more distant, and innate ability acquires many characteristics of merit.

To illustrate the widespread acceptance of the moral arbitrariness of innate ability that does not bring merit, we may quote Mises (1951) stating that 'talent and genius are the gifts of God, and the individual is not responsible for them, as is often said' and Hayek (2013) affirming that:

> Though either may greatly affect the value which an individual has for his fellows, no more credit belongs to him for having been born with desirable qualities than for having grown up under favourable circumstances.

It is unnecessary to quote egalitarians reaffirming the same point. But the difference between egalitarians and entitlement theorists arises when we come to the consequences of the unearned differences in innate ability.

Nearer the other end of the spectrum of philosophical opinion, we have Rawls, who famously saw innate ability to be a natural resource and thus a common asset, as 'there is no more reason to permit the distribution of income and wealth to be settled by the distribution of natural assets than by historical and social fortune.' Thus, the difference principle is constructed by Rawls (2009) so that we

> regard the distribution of natural talents as a common asset and to share in the benefits of this distribution whatever it turns out to be. Those who have been favoured by nature, whoever they are, may gain from their good fortune only on terms that improve the situation of those who have lost out.

The same view was expressed early on by Durkheim:

> is it any more favourable that he should be better treated because he was born of a father of higher intelligence or in a favourable moral milieu... [And] ... (e)ssentially are not . . . inequalities of merit fortuitous too? . . . man as scholar of genius or as an able man of business, and so on. . . For these [inequalities of merit] . . . all men are born with ... it hardly seems just to make them bear responsibility for them. (Durkheim as quoted by Green, 1989)

Entitlement theorists, in turn, have held the position that the moral arbitrariness of innate ability does not mean that the influence of innate ability on holdings is unjust.

> The physical fact that those particular gametes contained particular organized chemicals (the genes for people rather than for muskrats or trees) is arbitrary from a moral point of view; it is, from a moral point of view, an accident. […] the entitlement theory's claim that moral entitlements may arise from or be partially based upon such facts is what is now at issue. (Nozick, 2013)

Entitlement theorists hold that while innate ability may be morally arbitrary, it may create just entitlements to holdings. Here it may also be noted that the Rawlsian position is also more ambiguous. The role of an asset as the basis of placing individuals in the social system is logically independent

of the principles of sharing the benefits of those assets. Rawls (2009) writes that that in placement in social offices and positions,

> assuming that there is a distribution of natural assets, those who are at the same level of talent and ability, and have the same willingness to use them, should have the same prospects of success regardless of their initial place in the social system, that is, irrespective of the income class into which they are born.

This is a more substantive interpretation of the conception that careers should be open to all talents in that it adds background fairness to procedural justice. The choice of note, however, is including talent as a just cause of differences in competition for positions. As noted above, Rawlsian fair equality of opportunity matches Temkin's (2016) equal opportunity merit principle, which takes merit to refer to 'ability to perform up to the level required by the role or position in question', as both take innate ability to be a merit in the competition for positions based on the fact that it affects one's ability to perform in that position. So here merit excludes the dimension of being earned by choice or action.

The fact that Rawls (2009) sees talent as a just basis for selection into social positions introduces inner tension into his position when he holds that social and economic inequalities are to be arranged so that they 'are attached to offices and positions open to all under conditions of fair equality of opportunity.' Arneson (1999) justifiably notes that the attachment of social and economic inequalities to positions in a competition for which innate ability is a just cause of differential outcomes amounts to making innate ability a just cause for social and economic inequalities. Via Fair Equality of Opportunity:

> According to Rawls's principles of justice, possession of native talent entitles one to special advantages to which the untalented are not entitled, advantages which constrain the commitment of justice to helping the disadvantaged, under certain circumstances.

Arneson's (1999) reading of Rawls (2009) seems to be justified, but would consider that the inner tension in Rawls's position may be resolved simply by holding that innate ability may be a just cause for differences in offices and positions, because having the best ability in all positions benefits society at large, but not for social or economic inequalities. In this view, innate ability can be considered a legitimate cause for differential learning outcomes and social placement, independently of the question of whether innate ability justifies differential welfare.

Rejection of merit

Another tension to resolve presents itself to entitlement theorists: that between responsibility – based on freedom and thus linked to merit – and acceptance of innate ability as a just cause of social differences. For them, the recognition that innate ability does not bring merit any more than circumstances, because it is independent of individual action, also leads to rejection of the principle of distribution according to merit.

Why should this be? The key is the recognition by Hayek (2013) that 'To decide on merit […] presupposes also that we can distinguish between that part of their achievement which is due to circumstances within their control and that part which is not.' In this assertion he echoed von Mises, who also saw that assessing merit is intimately associated with assessing individual responsibility:

‘Similarly it is quite impossible to make the merit of the individual the general principle of distribution. Who is to decide on merits?’ (Mises, 1951)

Hayek concluded that the need to distinguish between circumstances and merit requires the ability to ‘judge in every individual how well people use the different opportunities and talents given to them’ and concludes that this puts distribution on the basis of merit in conflict with liberty, and led to his rejection of the principle of distribution on the basis of merit.

With distribution according to merit, equality of opportunity is also rejected.

> All human differences, whether they are differences in natural gifts or in opportunities, create unfair advantages. But, since the chief contribution of any individual is to make the best use of the accidents he encounters, success must to a great extent be a matter of chance. (Hayek, 2013)

In the same vein, Nozick explicitly rejects equality of opportunity. He considers that such equality may be provided ‘by directly worsening the situations of those more favoured with opportunity, or by improving the situation of those less well-favoured’, but any improvement in the position of some requires resources, and thus the improvement in the position of any ‘involves worsening the situation of some: those from whom holdings are taken in order to improve the situation of others.’ The holders of these assets are, however, entitled to them, and ‘no one has a right to something whose realization requires certain uses of things and activities that other people have rights and entitlements over.’ In short: ‘holdings to which these people are entitled may not be seized, even to provide equality of opportunity for others.’ (Nozick, 1974)

This rejection of equality of opportunity, with rejecting merit as the basis of distribution, provides grounds for the choice not to use the honorific ‘formal equality of opportunity’ for procedural justice, for which it is even more removed epithet than ‘equality before the law’. It sides with the analyses of von Mises, Hayek and Nozick that the entitlement theory approach to justice rejects both the principle of equality of opportunity and of distribution according to merit.

As an aside, it may also be noted that the rejection of merit sits in some contrast with the justification of responsibility by freedom and the need to accept outcomes as the consequences of one's actions, considered so fundamental to a free society. It is linked to a broader structural asymmetry in the normative position, as the raw facts of the world may bring one entitlements as well as rights that those entitlements are respected, but may not bring duties, as no one has a responsibility to rectify what Temkin would call natural injustices. This asymmetry in which brute facts have normative implications seems to sit at the heart of the balancing act that right-libertarianism and entitlement theory are engaged in. An extensive treatment would, unfortunately, take us rather far from the substance of this work.

Natural and socially conditioned consequences of innate ability

Both tensions within the Rawlsian and the entitlement theory positions are related to the interplay between natural necessities and moral justification of social arrangements. Rawls takes a natural distribution of innate ability as a given brute fact, and thus refuses to give it moral significance, but accepts it as a cause for allocating social positions. The tension between the positions arises from attaching social and economic inequalities to social positions allocated partially on the basis of innate ability.

In more radical interpretations, the idea of not holding individuals responsible for their innate ability leads to arguments that the 'equal opportunity merit principle' is not precisely a principle of equality of opportunity. It has been argued that

> it is not clear that the philosophical concept of equality of opportunity is the right term for this concept in education. Equality of opportunity implies holding people responsible for the things within their control and not for circumstances beyond their control; but the moral basis for holding children responsible for their own innate talent or motivation appears weak. For this reason, we refer to this type of equity concept in education as impartiality. (Cameron et al., 2018)

This solution tries to fend off the inequalities rising out of the unequal distribution of innate ability. As Thomas Nagel (1979) has put it:

> When racial and sexual injustice have been reduced, we shall still be left with the great injustice of the smart and the dumb, who are so differently rewarded for comparable effort.

Now, we may legitimately argue that welfare perhaps should, and social positions no doubt could, be distributed without concern for innate ability because it is morally arbitrary. But innate ability is, by definition, intimately causally linked to learning outcomes. Thus, while we may consider that, in the grand scheme of things, it is a failure of equal opportunity to be born with less innate ability than someone else, from a societal point of view, 'we would not normally regard meagre personal capacity as detracting from opportunity.' (Cohen, 2011)

Indeed, the challenge is how to fit together the inevitable, even conceptually necessary, link between innate ability and cognitive outcomes, the fact that they are morally arbitrary, and the need to justify the distribution of outcomes with moral arguments. For outcomes that are clearly socially determined, this is easier because we can recognise that the value of a trait comes from society, not from the individual. This has been widely recognised, even by entitlement theorists.

> It is good luck for him if market conditions are such that a kind of labour which he is able to perform is lavishly remunerated; it is chance, not personal merit if his innate talents are highly appreciated by his fellow men. Miss Greta Garbo, if she had lived a hundred years earlier, would probably have earned much less than she did in this age of moving pictures. (Mises, 2006)

But it also entails that, without additional assumptions, market outcomes (whatever the market in question) do not form a neutral baseline for distribution of outcomes. Outcome distribution is always due to the rules of interaction as well. Entitlement theorists tend to take the rules of the market, and thus the outcomes, as justified. Their critics argue that the question is about justifying a given set of rules. They cannot be justified by their consequent distribution if that distribution is in turn justified as one following from the rules. We may contend that, at least in theory, we may always imagine a world in which individual motivations and values would lead to allocation of welfare that would not follow from innate ability.

For the consequences of innate ability and learning outcomes, the matter is more difficult because learning outcomes, as well as innate ability, are inalienable and cannot be redistributed. Thus, we must find what we must take as a necessary consequence of natural differences, and what as a question of social choice. It may be noted that many social choices have been traditionally justified by natural inevitability:

> Just as the pseudo-democratic movement endeavours by decrees to efface natural and socially conditioned inequalities, just as it wants to make the strong equal to the weak, the talented to the untalented, and the healthy to the sick, so the radical wing of the women's movement seeks to make women the equal of men. (Mises, 1951)

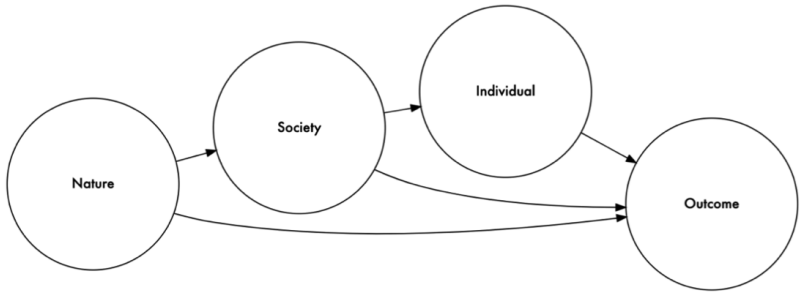
While some of these justifications from supposed biological inevitability now ring hollow, we are biological and material beings inexorably tied to the matter that constitutes us, or that we inhabit. The question is, where are the limits of nature and how do they fit the boundaries of normative judgement?

5.2 Responsibility without freedom

Causes of outcomes

Now, if innate ability is widely agreed to be morally arbitrary, but also widely agreed to be a justified cause of differences in learning outcomes, it is necessary to explicate how and why we may consider that to be the case. This is where we will use the distinction introduced above between principles of justice and rules of regulation. Our exercise in measuring inequality of opportunity is concerned with the consequences of unjust causes of outcomes. It amounts to making the cut between what individuals are held responsible for. At the same time, there are different types of causes, which are to varying degrees subject to individual or human influence.

Diagram 3: Determinants of outcomes



For our purposes, it is now enough to divide the possible determinants of individual outcomes into three broad categories of causes: individual, social and natural. The ‘individual’ refers to cases of causes where individual choice or action affect an outcome. The ‘social’ refers to causes of outcomes that are beyond the ability of an individual to bring about but are determined by official or unofficial social arrangements or by collective human action in general. ‘Natural’ is nature as something beyond even the collective influence or control of humans, the background to our existence.

This classification is a heuristic and covers all causes of outcomes in the universe. When it comes to outcomes of individuals belonging to – or living their lives in – a society, many, even most, outcomes are determined by a combination of the three. The social always acts within the limits set by the natural, and the individual within the limits set by the social and the natural. Thus, as our

interest here is in individual human outcomes, we do not usually seek outcomes that have been determined by one of the three, but rather for the contributions of different causes to the variation of outcomes.

Our interest lies in understanding how different outcomes are determined to understand the opportunity space available for individuals and societies. Understanding the causalities involved is crucial for policy, as policy needs effective tools to realise its objectives, but it is also crucial for normative assessment of the situation.

Responsibility and three determinants of outcomes

Only one of the causes, individual, is subject to individual choice, action and effort. This is the type of cause that both Temkinian canonical egalitarianism and Hayekian libertarianism hold individuals responsible for. The other two types of causes, social and natural, are morally arbitrary causes because they are not in any way under individual control or influence.

So far, so good. However, the responsibility parameter of a given normative system is by no means limited to holding individuals responsible only for causes within their control. It may be set to hold or not hold an *individual* morally responsible for any of the three types of determinants of outcomes. When we hold individuals responsible for social and natural causes, we simply hold them responsible, as a matter of justice, for causes that we do not deem them causally responsible for. Now, many of the common forms of equality of opportunity are based on taking it as unfair and unjust to hold individuals responsible for causes beyond their control, but things are not so simple. When we come to theories of justice that boil down to procedural justice, individuals may quite standardly be held responsible for causes beyond their control. (Hayek, 2012, 2013; Nozick, 1974)

The standard match between causal and moral responsibility is presented in ***Table 1.3a***, which presents a typology of moral and causal responsibility built on three different causes (individual, social and natural) and the result of the application of a responsibility parameter. The simple solution is to hold an individual morally responsible for any effort that is under her causal control and treat other causes as circumstances.

Table 1.3a: Effort and circumstances

		Causal responsibility (can)		
		Individual	Social	Natural
Moral responsibility (ought)	Yes (responsibility characteristics)	Normative responsibility characteristic (effort)		
	No (circumstances)		Circumstances	Circumstances

Three types of responsibility

For our current purposes, our main interest lies in the three empty cells of ***Table 1.3a*** where individuals are either held responsible for causes beyond their control or are not held responsible for things within their control. I classify responsibility characteristics into normative, historical and natural, referring to individuals being held responsible for individual, social and natural causes of outcomes.

While *normative responsibility characteristics* hold individuals responsible for their choices and effort, the two other types of responsibility characteristics involve individuals carrying a moral responsibility for outcomes or causes outside their causal influence. *Historical responsibility characteristics* refer to socially determined determinants of individual outcomes. These cover many aspects of social arrangement ranging from the highly abstract, like the choice of the *responsibility parameter,* to the practical. Social arrangements that affect individual outcomes belong in this category. *Natural responsibility characteristics* refer to characteristics that have causes beyond individual and collective human intervention and influence, but that individuals are held responsible for.

Causes of any of the three types may also be seen as not entailing individual moral responsibility. In the spheres of natural and social, they are then usually known as circumstances, as something beyond individual control and thus morally arbitrary. The more interesting class may be cases where an individual is causally responsible for an outcome but should not be held morally responsible for it. This will be referred to as *sufficiency* and refers to cases in which individual action would produce outcomes that are considered so poor that it is not just for the individual to bear them fully. The proposed concepts to fill the blanks of ***Table 1.3a*** are presented in ***Table 1.3b***.

Table 1.3b: Effort and circumstances

		Causal responsibility (can)		
		Individual	Social	Natural
Moral responsibility (ought)	Yes (responsibility characteristics)	Normative responsibility characteristic (effort)	Historical responsibility characteristic	Natural responsibility characteristic
	No (circumstances)	Sufficiency & proportionality	Circumstances	Circumstances

Fault and choice of one's own - Normative responsibility characteristic

A normative responsibility characteristic is the fundamental type of responsibility characteristic that is justified by the need for an individual to bear the consequences of their actions. Normative responsibility characteristics are all dependent on individual choice and effort and are thus responsive to incentives. Here 'the individual has both the opportunity and the burden of choice', but also bears 'the consequences of his actions' (Hayek, 2013), differences in outcomes are justified if they have arisen from the 'fault or choice' of individuals (Temkin, 1993) and *do* 'appropriately reflect choices that he has made or is making or would make.' (Cohen, 2011) In short, normative responsibility characteristics are those in which moral and causal responsibility meet. As this link between freedom and responsibility has been already discussed above, no more need be said here.

Nature as just - Natural responsibility characteristics

Sometimes we encounter normatively pertinent brute facts. These facts are referred to here as *natural responsibility characteristics*. The need for the concept arises from the fact that one unavoidable cause of learning outcomes is innate ability. While it is quite obviously beyond individual choice and thus morally arbitrary, it cannot be redistributed. Thus, there are good reasons to – and not to – consider it unjust, with most authors implicitly or explicitly coming down on the side of considering it to be a legitimate cause of cognitive (or other) outcomes.

Natural responsibility characteristics need to fulfil at least three criteria to be considered as such: 1) immutability, 2) inalienability, and 3) outcome specificity. Together these three criteria put the natural responsibility characteristic beyond human intervention.

Immutability requires that the characteristic cannot be changed by human intervention. This essentially defines potential natural responsibility characteristics as being outside all human influence.

It is to be noted that here immutability does not mean complete unresponsiveness to influence. A characteristic is considered immutable if the characteristic cannot be altered in a direction that would be linked with better outcomes. This is a critical proviso as only such cases are ruled out where equality is achieved by means of levelling down. Many congenital conditions may significantly limit possible cognitive and physiological outcomes and be beyond cure. The distribution of such outcomes in the population might be equalised if the same condition were caused to the otherwise unafflicted or their circumstances changed in such a way as to lead their outcomes to be as if they were afflicted. However, 'levelling down seems a waste, pure and simple, and everyone's having the same does not seem in any way intrinsically worthwhile'. (Arneson, 2000) Thus, inequality is just because equality would lead to levelling down.

Inalienability requires that the characteristic cannot be redistributed between individuals. If it could, we could make a choice. Here we are in the area of natural primary goods (Rawls, 2009), more specifically, those that are non-transferable resources. (Green, 1989)

Outcome specificity requires that the characteristic be specific to certain outcomes by a causal chain leading from the characteristic to the outcome, which cannot be severed or modified. While immutability requires that the characteristic is not mutable, this requirement pertains to the immutability of the relevant causal consequences of a characteristic. Innate ability is not a just cause of differential learning outcomes if learning outcomes were causally independent from innate ability. Moreover, while innate ability might be a natural responsibility characteristic for learning outcomes, it is not necessarily one for welfare as the relationship between learning outcomes and welfare can be modified.

Together the three criteria fix natural responsibility characteristics as unavoidable determinants, however loose, of the distribution of possible outcomes of an individual which are beyond individual or collective human action.

Natural responsibility characteristics are of interest to those who are interested in equality of learning opportunity, or equality of opportunity for achieving an outcome that is not purely socially determined. Learning outcomes are partially dependent on innate ability; this innate ability is not uniformly distributed and is clearly outside individual control. If we interpret this to entail that innate ability is a circumstance, then differences in innate ability are unjust and must be equalised for justice to prevail. As innate ability is not directly distributable, redistribution could only be addressed by

destroying part of the innate ability of the more innately able, which would open our reasoning to the levelling-down objection. Thus, innate ability will be considered a natural responsibility characteristic for learning outcomes.

Attention may be drawn to the fact that normative responsibility characteristics fall squarely in the area of rules of regulation. The immutability of such characteristics is historically contingent, i.e. dependent on the level of technology available, and thus any natural responsibility characteristic is one only provisionally. If we enter a brave new world in which the distribution of innate ability is subject to human influence, it becomes subject to different normative considerations. (Huxley, 2007) In terms of fundamental fact-insensitive principles of justice, we may quite reasonably consider that the inequality of natural responsibility characteristics, including inequality of innate ability, is unjust. If it were possible to address this inequality through policy intervention, it would be very difficult to morally justify a distribution of innate ability that would match the current one.

Justice in process - Historical responsibility characteristics

Historical responsibility characteristics sit between natural responsibility characteristics, which set the background for all of humankind, and normative responsibility characteristics, where we face the world as morally responsible individuals. This class of determinants of outcomes lies beyond individual control, but not beyond the control of humans collectively.

The essential difference between natural and historical responsibility characteristics is the difference between nature and the *socially determined consequences* of nature. It may be accepted that earthquakes are beyond our control, without accepting that individuals who have suffered from the consequences of one should not receive help in their situation. By the same token, we may well accept that congenital conditions affect the cognitive capacity or physical functioning of an individual in a way that is unresponsive to intervention and that such differences are to be accepted as inevitable. This position does not amount to accepting, nor does it make it necessary to accept, that individuals with lower levels of innate physical or cognitive ability should suffer lower welfare.

While natural responsibility characteristics must be natural in the sense of being unavoidable, historical ones may be currently unavoidable for an individual but might be avoidable under a different social arrangement. Thus, historical responsibility characteristics can be required to be justified by a normative justification in a way that natural ones cannot. While social facts may be even more brutal than natural ones, they are less brute.

Lightening the unbearable – Sufficiency and proportionality

While equality of opportunity is a crucial equalisandum, it is not clear that as long as opportunities are equal, any distribution of outcomes that follows is just. There are two relevant considerations – the form of the reward function and the level of the lowest outcomes.

Very unequal distributions of outcomes also raise the possibility that some outcomes are so low that they can be considered unjust. Here we leave the area of equality and enter that of sufficiency, that each should have enough, argued for by Frankfurt (2015) against equality as a normative ideal. The sufficiency doctrine is not concerned with equality of the distribution or with how it came about, but solely on whether all individuals have enough.

But considerations of sufficiency are not the only aspect of injustice that may go beyond equality of opportunity. The principle of responsibility does not automatically entail any single *reward principle,* and thus the same scale of responsibility characteristics may be associated with outcome differences that range from the tiny to the gargantuan. In this sense, equality of opportunity is interested in ordinal rankings and assesses the ordinal ranking of responsibility characteristics against the ordinal ranking of outcomes. Equality of opportunity fails to obtain if the two ordinal rankings do not match. But as there is no commensurate way to measure responsibility and outcomes, the distribution of responsibility characteristics implies nothing in terms of the cardinal distribution of outcomes (Fleurbaey, 2008). Thus, 'considerations outside the realm of equality of opportunity must be brought to bear to decide upon how much inequality with respect to differential effort is allowable'. (Roemer, 1998) As long as the shape of the reward function is considered over the entire distribution, the question can also be understood as not being about the relationship between effort and mercy as distributional principles, but about the application of the reward principle.

The schema

These considerations are directly relevant to criticisms of egalitarianism as a revolt against nature. (Rothbard, 2000) Normativity can revolt against nature only if nature has normative implications. On the level of rules of regulation, revolting against nature is futile: no one can have a duty to overturn natural inevitability and thus natural inevitabilities are as good as they are just. It is less clear that this applies on the level of principles of justice. Accepting the emerging outcome as just requires accepting the principle that what emerges from the existing process in the market or in evolutionary competition is just.

The solution offered here cuts the Gordian knot by including in the class of normative responsibility characteristics all characteristics that can be justified by normative arguments, whereby emergent outcomes can be just if the process producing them can be so justified. When a purely normative argument is not present, we accept some outcomes as just, by natural or historical inevitability. To the degree that we – or any conscious moral agents – approach omnipotence, the scope of historical and natural responsibility characteristics dwindles. Some may become circumstances when they become malleable; some may find normative arguments to support them and become acceptable as normative responsibility characteristics. Here, we do not need to guess how that process plays out.

5.3 What is measured

Making the empirical cut between equality of opportunity and equality of opportunity

On the whole, internal resources such as innate ability are not readily amenable to redistribution, and may very legitimately be the basis of placement of individuals in a social system irrespective of whether their ability brings them any benefits through this placement. The problem, of course, remains: innate ability remains outside individual control even if is a legitimate basis for placement in social structures.

While the above criticism that impartiality holds individuals responsible for characteristics beyond their control, in empirical work impartiality is the dominant criterion of inequality of opportunity, despite the fact that it holds individuals (partially) responsible for their innate ability. Thus, it is useful to clarify what normative positions may be measured by different empirical approaches to equality of opportunity, considering that all the empirical strategies are plagued by the fact that all of the core variables at play are unobservable.

This is an important distinction as, if it is accepted that individuals differ in innate ability, then insisting on equality of learning outcomes irrespective of the level of innate ability would make the position of any level at which full equality could be achieved very low indeed. Here it is considered that, in terms of learning outcomes, equality of opportunity should be understood as impartiality, without this choice implying anything about the choice of principle for treatment of innate ability in the distribution of welfare in society.

The available positions have been summarised in ***Table 1.4a***. It also allows the reader to note that, in terms of holding individuals responsible for their different characteristics, the conception of equality of opportunity as impartiality to be used in the empirical sections of this work is closer to entitlement theory than to equality of outcomes. The extremes of the axis of principles of equality may be understood as overcompensation in response to perceived risks of misidentification of effort or circumstances. Equality of outcomes is prepared to distribute the results of effort evenly to avoid the injustice of circumstances dictating the outcomes. Entitlement theory is ready to hold individuals responsible for things beyond their control to avoid having them escape responsibility for things that they can be held responsible for.

Equality of opportunity as *ability as common asset* occupies the exact middle ground, as it would take innate ability, which is clearly beyond individual control, among the circumstances whose effect on distribution is to be neutralised. For this purpose, it needs accurate information on which parts of the distribution emanate from circumstances and which from the actions of the individuals in those circumstances (effort).

Equality of opportunity as *impartiality* is less prepared to adopt the high-handed philosophical position of treating innate ability as a natural resource to be used for the common good. When choosing whether it should err, from the point of perfect information, on the side of equalising outcomes (as equality of outcomes) or holding individuals responsible for their circumstances (as entitlement theory), equality of opportunity as impartiality chooses to err on the side of responsibility and holds individuals responsible for their innate ability as well.

Table 1.4a: Innate ability and equality of opportunity

	Circumstances	**Ability**	**Effort**
Equality of outcomes	Equal	Equal	Equal
Equality of opportunity (ability as a common asset)	Equal	Equal	Responsibility
Equality of opportunity (impartiality)	Equal	Responsibility	Responsibility
Entitlement theory	Responsibility	Responsibility	Responsibility

In ***Table 1.4a***, we still assume that ability is independent of circumstances. As mentioned above, this assumption becomes debatable when innate characteristics can be responsibility characteristics. Which brings us to the fact that, depending on the relative contributions of innate ability and circumstances to outcomes, the between-type equality of opportunity, the very same empirical measure ends up serving very different normative commitments.

This can be illustrated by turning ***Table 1.4a*** into ***1.4b*** with some additional details. The new table makes two new distinctions. Firstly, it distinguishes two different types of innate ability, one particular to the individual (varying within type), the other associated with her background (varying between types). As can be seen, this new distinction complicates matters for responsibility sensitive egalitarianism, while equality of outcomes and entitlement theory are not affected.

The crucial issue is that, by making a distinction between two types of innate ability, the table allows us to see that empirical strategies that cannot observe innate ability – that is, all of them – cannot distinguish empirically between ability as a common asset and impartiality. There is a corresponding empirical identification problem that becomes visible when impartiality is divided into two versions. The added new position, referred to as equality of opportunity as *circumstance impartiality*, also holds individuals responsible for the part of their individual talent that is associated with their background. Thus, innate ability awards individuals rights to rewards even if it is a group characteristic beyond individual control.

The old (empirical) version of impartiality, being by practical necessity blind to any covariation of individual ability with socioeconomic background, holds individuals responsible for their ability that is independent of their background, while inequalities due to socioeconomically differential ability are indistinguishable from any other inequalities due to background. The added distinctions in the table make it easier to appreciate that, while empirical strategies are sufficient for the top, middle and bottom rows, the second and fourth rows hold greater normative appeal. It is intuitively more natural to hold individuals responsible for all of their innate ability or none of it, than for some of it.

The added detail also illustrates how empirical measures of *impartiality* end up approximating *ability as a common asset* or *circumstance impartiality,* depending on how strongly innate ability is associated with background. If type is determined by ability, empirical *impartiality* effectively

Table 1.4b: Innate ability and equality of opportunity

	Circumstances	**Innate ability (associated with background)**	**Innate ability (independent of background)**	**Effort**
Equality of outcomes	Equal	Equal	Equal	Equal
Equality of opportunity (ability as a common asset)	Equal	Equal	Equal	Responsibility
Equality of opportunity (impartiality)	Equal	Equal	Responsibility	Responsibility
Equality of opportunity (circumstance impartiality)	Equal	Responsibility	Responsibility	Responsibility
Entitlement theory	Responsibility	Responsibility	Responsibility	Responsibility

measures *ability as a common asset*. Correspondingly, if innate ability is independent of type, *impartiality* effectively measures *circumstance impartiality*. Thus, the nature of the measure depends on how an unobservable variable, innate ability, is distributed as a function of circumstances. Or rather, whether circumstances are distributed as a function of innate ability or not.

Thus, while in the following we will use standard between-types equality of opportunity as an empirical measure of equality of opportunity, matching impartiality in the above table, it is done as a second-best option. We would prefer to measure circumstance impartiality, so that we could hold individuals responsible, for learning outcomes, and for their innate ability. As noted, this choice does not constrain the choices of possible positions on equality of opportunity for welfare, and leaves a range of positions available, from Rawlsian ability as a common asset to very laissez-faire approaches that make individuals completely responsible for their innate ability.[15]

[15] While selection of right principles of equality of opportunity for welfare are important for any society, this work does not need to be concerned with any of the relevant intricacies and can thus leave them untouched.

SECTION III – GROUPS, TRADE-OFFS AND WHAT COULD BE

6 Ability, opportunity and the equity-efficiency trade-off

Can we get quality and equality, or do we have to choose? Well, that depends on the type of outcomes and the type of equality we want. The question is of obvious importance in policy, as both the level of outcomes and the realisation of social equality are important goals in all policy areas. It is very important to know if both can be achieved simultaneously, with the pursuit of one not deducting from the other, or if progress in one has to be bought by accepting lower achievement of the other. In education, Pfeffer (2015) has recognised that the 'primary concern of educational policy-making is whether socio-economic equality in educational opportunities can be increased without lowering the quality of education.'

The equality-efficiency trade-off, as we know it today, first rose in economics as a trade-off between efficiency and economic growth, before Okun (2015) generalised the concept into a general quality-equality or equality-efficiency trade-off between the level and equality of distribution of outcomes. From economics, the trade-off has spread to other fields of study, and in education sociology, it has been recognised that it is a fundamental question whether socio-economic equality in educational opportunities can be increased without lowering the quality of education. For our purposes, it is imperative to appreciate the fact that the sociological question above is fundamentally different from Okun's trade-off.

6.1 Trading off

The canonical outcome trade-off

Economic research and policy of income distribution have often been concerned with equality of outcomes and the trade-off of equality with efficiency and economic growth. (Kuznets, 1955) This trade-off was generalised by Okun (2015) into a general trade-off between the level and equality of distribution of outcomes, henceforth referred to as the *outcome trade-off*.

The outcome trade-off arises due to the harmful effects of equality of outcomes on incentives. Incentives are dampened through two main channels. The equalisation of outcomes has been seen to affect work incentives as the payoff for increased effort decreases. Relatedly, but through a different mechanism, equalisation of outcomes has been seen to affect incentives to save and invest, and thus reduce capital formation. Or, as Kenworthy (1995) sums up: 'Equality, in effect, crowds out investment.' The outcome trade-off arises, for all intents and purposes, from the diminished role of individual responsibility. The equalisation of outcomes dampens incentives to act in a manner that would yield higher levels of outcomes, which leads to lower output quantity or quality.

Outcome trade-off may also arise if we seek to equalise holdings in goods that are not redistributable. This type of outcome trade-off is more specific, as the emergence of the trade-off depends on the nature of the equalised outcome: it arises if there are innate differences between individuals in the level of attainable outcome. For example, in learning outcomes, the only way to achieve full equality of outcomes would be to achieve it by levelling down to a fairly low level, because the level would be determined by the individuals with low innate ability. The impossibility of redistribution

and the avoidance of levelling down give reasonable grounds not to consider equality of outcomes as a value in the area of cognitive outcomes. This applies at least to the degree that the attainable level of cognitive outcomes is limited by the brute fact of innate ability.

Equality-efficiency trade-off is closely linked to the levelling-down objection in philosophic settings, which objects to equalisation efforts that lead to sacrifices in overall levels of attainment. It is widely recognised that equalisation of any equalisandum by levelling down attained levels of achievement is problematic as 'levelling down seems a waste, pure and simple, and everyone's having the same does not seem in any way intrinsically worthwhile.' (Arneson, 2000)[16] The outcome trade-off provides a mechanism for how, in certain circumstances, levelling of outcomes happens down instead of up.

Lack of opportunity as inefficiency

The demands of equality of outcomes are different from those of equality of opportunity, which renders individuals responsible for their choices or effort. Okun (2015) highlighted already in his seminal work the difference between equality of outcome and equality of opportunity, noting that equality of opportunity is not in a trade-off with efficiency, but contributes positively to it. The reasoning is simple; inequality of opportunity entails that opportunities/resources are allocated to different individuals than they would be under competition on a level playing field. Individuals receiving resources that they would not get in a fair competition are, by definition, lower in terms of ability than the individuals who would get the resources in fair competition, and thus resources are inefficiently allocated.

> if women are excluded from responsible jobs, they are prevented from using their skills to the fullest extent; that is inefficiency […] Moreover, unequal opportunity at one point in time generates unequal opportunity over time. Once people are excluded from good jobs, they are deprived of the incentives and opportunities to develop the skills that would otherwise qualify them for good jobs. […] Thus, inefficiency can grow at compound interest. (Okun, 2015)

The marked contrast between equality of opportunity and equality of outcomes in this respect is the different effect of the distributions of outcomes and opportunities for incentives and opportunities. While an equal distribution of outcomes may deprive individuals of incentives to develop their skills, equal distribution of opportunities does not have the same effect. In fact, it has the opposite effect. As Okun argues, inequality of opportunity may present a source of inefficiency, as some of the natural, innate abilities of the population would be left underutilized due to the unequal distribution of opportunity.

This observation is linked to the difference between raw and accountable effort used in measuring inequality of opportunity, in which distributions of accountable effort are the same in all types even if distributions of measured effort might differ. The lower level of raw effort expended by a type

[16] The levelling-down objection is here accepted as a conclusive argument for holding individuals responsible for their innate ability even if it is beyond their control. Temkin has presented a thoughtful criticism of the levelling-down objection, showing that while the objection may have analytical force, it is by no means the trump card that it has been taken to be. He convincingly argues that proportional justice is most plausibly interpreted to be an impersonal principle of justice and points out that, according to proportional justice, we may judge a situation to be better when faring well matches doing well (holdings match merits) even if no one would fare better in that situation. (Temkin, 2000)

in less favourable circumstances is simply part of the mechanism by which unequal distribution of opportunity leads to efficiency loss.

The opportunity trade-off

Pfeffer (2015) sees that the

> fundamental question in sociological research on education and a primary concern of educational policy-making is whether socio-economic equality in educational opportunities can be increased without lowering the quality of education. I label this the potential equality–quality trade-off in education.

It needs to be noted that this trade-off is fundamentally different from the outcomes trade-off because equality is not measured in terms of outcomes but with a bivariate distribution of outcomes against background characteristics. In other words: in terms of opportunity. Thus, this trade-off will be referred to as *opportunity trade-off,* to distinguish it from the *outcome trade-off*.

The crucial question is, why should there be a trade-off between the socioeconomic equality of educational opportunity and quality of education when equal opportunity should, in general, increase efficiency rather than be in a trade-off with it? The equal distribution of opportunity, or redistribution of opportunity, does not hamper incentives, as equality of opportunity holds individuals responsible for their effort, and thus differences in effort are legitimate grounds for differences in educational outcomes. Equality of opportunity is uninterested in the equality of the outcome distribution. It allows any amount of inequality due to differential innate ability and is only concerned with differences in achievement attributable to circumstances.

Indeed, opportunity trade-off arises when there are between-type differences in innate ability. The interconnectedness of innate ability and socioeconomic background is necessary to explain why the realisation of greater equality of opportunity would lead to lower outcomes, given that equality of opportunity only requires that outcomes due to circumstances are equalised. If innate ability is independent of circumstances, equality of opportunity cannot lead to equalisation of outcomes flowing from innate ability, as all within-type differences are removed when between-types inequality of opportunity is measured. But if innate ability is associated with socioeconomic background, equalisation of opportunity will lead to the equalisation of outcomes not due to differences in opportunities but to between-types differences in innate ability.

When it is recognised that the *opportunity trade-off* operates through the distribution of innate ability in the population, it is easy to see that the opportunity trade-off is not a recently identified trade-off at all. After all, the opportunity trade-off presents the view that individuals differ in innate ability, the best among the populace constituting a *natural aristocracy,* and combines it with a contention that the natural aristocracy is at least partially reflected in the *observed aristocracy* of birth. The view that abilities are distributed in different social groups so as to give each group abilities that correspond to its position in society, or a position that corresponds to its particular abilities and talents, has long historical roots.

Typology of trade-offs

The outcome trade-off arises through effects on the effort distribution channel, while the negative association between inequality and outcomes operates through the opportunity channel. The causes for the emergence of trade-offs are summarised in ***Table 1.5***, distinguishing outcome trade-off from opportunity trade-off and effort from innate characteristics. The difference between outcome and opportunity trade-offs is the focus on individual and group differences, and opportunity trade-off may only arise out of group differences. Because measures of equality of opportunity usually treat group differences in effort as functions of the different circumstances of groups, opportunity trade-off does not emerge for effort, but only for innate characteristics.

Table 1.5 - Determinants of equality-efficiency trade-offs

	Effort	Innate characteristics
Outcome trade-off	Individual differences in effort	Innate differences between individuals
Opportunity trade-off	Group differences in effort	Innate differences between groups

6.2 Ability and opportunity hypotheses

Ability and opportunity hypotheses

This is where we come to the ability and opportunity hypotheses. In the presentation above, innate ability has been presented in the context of individual responsibility, and some observations were made on the relationship between empirical measures of equality of opportunity and the distribution of innate ability relative to circumstances to note that the normative constructs measured by common empirical measures depend on the level of association between background and innate ability. In the section on trade-offs, we noted that, as common measures of equality of opportunity are interested in between-type differences, innate ability acts as a confounder of the effect of circumstances on outcomes only if types defined by circumstance characteristics are selected by innate ability. We also noted that the fundamental question of education policy is thus essentially the age-old question of whether or not social positions have been determined by the characteristics of the individual or by favourable or unfavourable circumstances. For between-types inequality of opportunity, the question is: do groups differ by innate ability so that inequalities of ability get misidentified as inequalities of opportunity? This question is not novel at all, but rather one that has followed us, if not from the caves of the Palaeolithic, at least from the dawn of recorded civilisation.

Innate ability as potential

Many of the complexities involved in measuring equality of learning opportunity are related to the role of innate ability. Unfortunately, even if the achieved ability can be measured, innate ability is

not directly observable any more than opportunity is, as it similarly exists only in potential, and the potential must be inferred from what is observed.

Before going into greater detail, it may be noted that, as all innate tendencies are always realised in interaction with the environment, even inheritable innate resources are not environment-invariant. Genetically determined potential is not realised irrespective of circumstances. As realised outcomes are a function of innate ability and the environment or opportunity, any measured outcome may arise due to very different interplays of innate ability and opportunity.

The concept of innateness has been subject to much criticism, even having been characterised as emerging from folk biological essentialism, and referring to several possibly related, but nonetheless biological properties. The term has encompassed developmental fixity, as unresponsiveness to environmental changes or differences; species nature, as a typical or universal property of a species; and intended outcome, as the meant outcome of individual development. Conflation of these different biological properties into one term has been interpreted as rendering the concept hopelessly confused. After all, a trait may be innate in being explained by biological rather than environmental factors, but not innate in the sense of being unchanging through development, shared by all members of the species, or being present at birth. (Griffiths, 2002) While the term has also been defended (Khalidi, 2007), the concept can be considered to lead to confusion beyond evolutionary and developmental biology, because it so easily matches our folk biological intuitions. Griffiths's advice is to forego using the term innate and simply refer directly to whatever properties are meant.

Throughout this work, the term innate ability is used to refer to the biological development potential of an individual. Here, the author is not heeding Griffiths's advice. The choice of terminology emphasises talent or biological potential, the difference between the aspects of ability that are within and beyond human influence, either individually or collectively. This biological potential is sometimes referred to simply as ability or intelligence, which might lead to the misinterpretation that observed abilities or intelligence express a permanent, essential characteristic of an individual. The reference to innateness is included in the term to make clear that we are referring to an innate characteristic, not an observed level of ability, which also depends on the environment.

In the following, innate ability is conceived as cognitive potential, understood as the level of cognitive ability that an individual would achieve under optimal circumstances, given her biology. For the purposes of this presentation, it is not necessary to specify how high the achieved outcomes would be if the potential under standard or optimal conditions were achieved, nor what would constitute optimal or standard circumstances. Nor is it necessary to assume that the same set of circumstances would be optimal for all individuals. What is optimal for one individual may not be optimal for another, and what is optimal for the development of one set of abilities might not be optimal for the development of another set of abilities in the same individual. Here optimal refers to optimal in the sense of maximising any set of abilities that determine social position.

Research on the acquisition of expertise has demonstrated that the limits of human potential may in some respects lie further than has been sometimes appreciated (Ericsson & Pool, 2016), but the cognitive potential of biological humans is not limitless, even in the best of circumstances. Environment affects the realisation of ability, and it does so within the limits posed by biology.

This conceptualisation of innate ability is not only unobservable but is, in fact, hypothetical, being brought to existence by its counterfactual consequences. The intuitive measure of such innate ability would be conceived of as the maximum level of cognitive performance achievable, in any

environment, with a given genotype. We run into some challenges as any characteristics can be valuable, and thus to be maximised, as a function of the environment, so that environment-invariant maximal performance is not possible even conceptually. (Lewontin, 1996)

In short, innate ability is a counterfactual concept that describes what could be, not what is. As such, it is a perfect pair for equality of opportunity, which is also about counterfactuals. The challenge for both is that we only have what is to try to figure out what could be.

Individuals and groups – malleability and selection

Life outcomes may depend on little other than our initial endowments and our environments. Across history, different emphasis has been placed on the relative contribution of each to how we turn out. Before going to an overview of what are here referred to as the ability and opportunity hypotheses, which emphasise natural endowments and environment respectively, it is useful to say a few words on the somewhat different shades of the question on individual and group levels.

On the individual level, the question of the relative contributions of environment and natural endowments is predominantly one of malleability. If humans were blank slates that could turn into anything, the environment would be all-powerful. At the other extreme would be the view that no change in the environment could affect how our outcomes turn out. While the range of positions is definitely extended, and many behaviourists have taken humans to be very malleable, it is difficult to find examples in the literature of cases where humans would have been taken to be blank slates, infinitely malleable and uniform in their limitless cognitive potential. For example, many (e.g. Pinker, 2004) quote Watson (1924):

> Give me a dozen healthy infants, well-formed, and my own specified world to bring them up in and I'll guarantee to take any one at random and train him to become any type of specialist I might select — doctor, lawyer, artist, merchant-chief, and yes, even beggar-man and thief, regardless of his talents, penchants, tendencies, abilities, vocations, and race of his ancestors.

What is often left unquoted, is his following statement that:

> I am going beyond my facts and I admit it, but so have the advocates of the contrary and they have been doing so for many thousands of years.

Here the focus is not on establishing who said what and when. It is sufficient to note that it is not uncommon to find in discussions on the biological or social origins of outcomes that contrasting views are quoted in a way that gives an exaggerated picture of the position defended, which is unfortunate and unhelpful. The exaggerated positions can be very useful in setting the scene in the same way that equality of outcomes is a useful concept, even if no one has ever advocated for the sameness of all outcomes.

On the group level, malleability is joined by selection as a possible cause of outcomes. The emergence of between-group differences in innate ability requires that outcomes not be entirely environmentally malleable and that the non-malleable traits affecting outcomes can be selected for. Even if individual differences in outcomes were almost exclusively determined by initial endowments, with the environment playing a minor role, group differences in outcomes would only be

explained by endowment if there were between-group selection in endowment. The difference between the two levels should be noted because of this difference in possible causes of differences in outcomes.

Ability and Opportunity hypotheses

In their pure forms, the two hypotheses can be presented as in ***Diagrams 1e*** and ***1f***, with the opportunity hypothesis seeing socioeconomic status as a mediator of circumstances, while ability hypothesis sees it as a mediator of innate ability. It is good to note that only the aspects of the hypotheses relevant for the causal path mediated by type are concerned here. Within-type variation in learning outcomes may be explained by both circumstances and innate ability, but once within-type variation is removed, only between-type variation in circumstances or innate ability remains relevant.

The stylisation of both hypotheses is somewhat extreme, as it posits that type is determined only by environmental or innate causes. On the group level the range of positions is extensive, from those interpreting that current evidence shows group differences to be primarily environmental (Nisbett et al., 2012a, 2012b) to those taking group differences as primarily biological. (Eysenck, 1998; Murray & Herrnstein, 1994; Gottfredson, 2005)

The sometimes ambivalent relationship between the individual and group levels has possibly helped the efforts of all parties of the debate to present themselves as occupying a moderate position that emphasises both environmental and genetic influences, and the other side defending an indefensibly strong form of genetic/environmental determinism. Hereditarians contrast the 'culture-only' model, where 100 % of the observed cognitive differences between blacks and whites are environmental, with a 'hereditarian' model that has '50 % genetic – 50 % environmental' determinants (Rushton & Jensen, 2005), and many writers refer to a supposedly widespread blank slate hypothesis. (Haier, 2016; Pinker, 2004)

In treatment terms, the ability hypothesis posits that the ill-defined treatment of group membership (social class, race, gender) is predominantly or exclusively well-defined genetic treatment, while opportunity hypothesis sees the handle of the well-defined treatment on the side of the environment. The difference is considerable from a policy perspective, because 'Changing behaviour by changing parental attitudes is a decidedly different proposition than tinkering with the ribosomes.' (Plomin, DeFries, & Loehlin, 1977)

The reason for referring to groups in general in this context, rather than focusing on socioeconomic groups or, more specifically, on groups defined by educational attainment, is twofold. First, while the question is immediately relevant to our purposes only in terms of parental education, the same considerations have been present in discussions on group differences along several dimensions, most prominent among them sex, ethnic group or race and socioeconomic position. Different groupings of individuals are candidates for different reasons, Some, because of their separate evolutionary histories, may have led to biological differences between them, others because of the selection mechanisms involved. While successful individuals do not change sex or race when they succeed, they may change their class status. Thus, the existence or absence of group differences in one classification does not necessarily imply anything in terms of their existence when groups are defined by other criteria. Class differences may exist despite the absence of racial or race differences. Or vice versa. What unites the discussions is the difference of opinions on whether we can take observed differences

in outcomes as indicative of deeper, permanent characteristics of the groups or in the social and environmental setting in which the groups find themselves.

Second, the discussions overlap factually, as socioeconomic and racial differences have overlapped in the US and in many other countries in which the discussion has been active. Because our interest in the matter is contemporary rather than historical, it is sufficient to note that von Mises, for example, explained differences on all three axes using innate differences (Mises, 1951) and that more contemporarily Gottfredson has linked both racial and socioeconomic differences in outcomes to innate differences in cognitive capacity. (Gottfredson, 2000, 2004, 2005, 2013) The intertwining of the different questions is of interest, because if e.g. class and race differences co-occur, it may be that race differences are due to discrimination, but class differences due to impartial selection. (Saunders, 1996, 2010) However, the existence of group differences due to discrimination in one respect means that the impartiality of operative selection mechanisms in other respects cannot be freely assumed.

The role of sex as a dimension of the discussion on biological, cognitive inequality has diminished as gender equality has progressed, and as the advantage of males over females in schooling and cognitive achievement has been reversed in very many areas in the majority of developed countries.

Selection for ability

Because the *opportunity trade-off* is only relevant when there is between-type selection by innate ability, our interest is focused on the question of selection on the group level, rather than on the question of responsiveness of learning outcomes to environmental intervention.

It can also be noted that the controversy and disagreement in the discussion has been, by and large, about the determination of outcomes on the group level, rather than on the individual level. It seems to be widely appreciated that individuals differ biologically or have different opportunities presented to them and that both of these factors play a role in how their lives turn out. It is less clear that certain groups are defined by characteristics

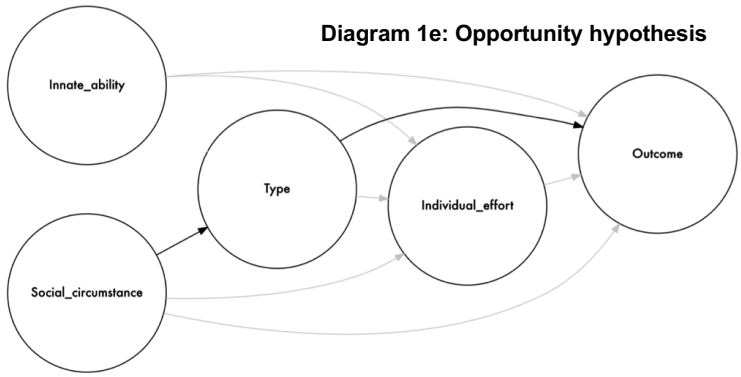


Diagram 1e: Opportunity hypothesis

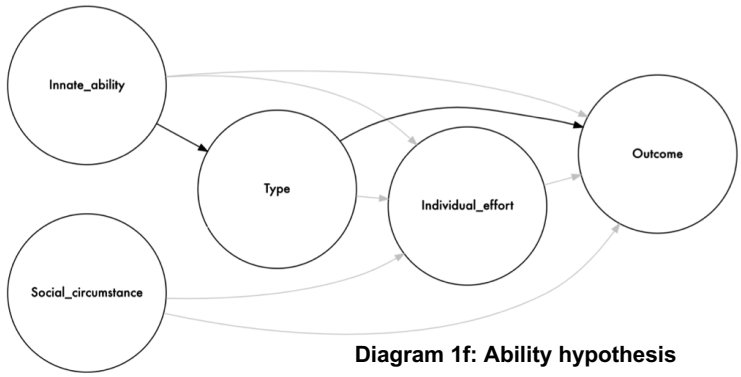


Diagram 1f: Ability hypothesis

that are associated with biological causes which make their life outcomes different from those of another group. The individual and the group levels are sometimes conflated in the discussion, which leads to less than charitable presentations of the contrasting points of view.

For the group level, the critical question is not whether innate ability exists or is hereditary. Both can be accepted without the question of group differences being resolved. The key question is whether there are mechanisms of selection that make groups differ by their innate ability.

The *ability hypothesis* posits that existing selection mechanisms have already produced a population stratified by innate ability. Thus, the persistence of positions across generations is explained by between-type differences in innate ability, and the lower achievement of lower socioeconomic strata reflects their lower levels of non-malleable innate ability. The fundamental presuppositions of this state of affairs were succinctly described by Herrnstein (1973), one of the foremost hereditarians of the late 1900s:

> (1) If differences in mental abilities are inherited, and (2) if success requires those abilities, and (3) if earnings and prestige depend on success, (4) then social standing (which reflects earnings and prestige) will be based to some extent on inherited differences among people.

Thus, the core presupposition of the ability hypothesis is that selection mechanisms can continuously select for innate ability even if environmental factors place a thumb on the scales by affecting the expression of inborn traits. It recognises that the selection mechanism needs continuous movement between social groups so that Youngian meritocracy might prevail. Jensen (2012) presents the relevant forces at play:

> In fact, if social classes rigidified into castes at some period in history, genetic intelligence differences between them would most likely be reduced, since all of the IQ variability arising within classes in each generation would remain as within-class variance. A high degree of social mobility correlated with ability, on the other hand, in each generation 'converts' a substantial proportion of the within-class variance to between-class variance. Thus, classes separated by more than two or three steps in the SES hierarchy can in time undergo wide separation in the distributions of genetic factors related to ability. This trend increases the closer we approach equality of educational and occupational opportunity and the more that SES mobility reflects ability factors rather than inequalities in opportunity.

The relevant question is whether the conditions for the emergence of between-group differences have been present and whether there has been enough time for the relevant forces to play out and achieve a match between natural and social inequalities. Herrnstein (1971) clearly thought that

> …we have been living with an inherited stratification of our society for some time. […] The opportunity for social mobility across classes assures the biological distinctiveness of each class, for the unusual offspring—whether more or less able than his or her) closest relatives—would quickly rise above his family or sink below it, and take his place, both biologically and socially, with his peers.

It may be noted here that the ability hypothesis is linked with taking educational attainment to be purely positional. While educational attainment can be positional for social reasons as well, ability hypothesis would lead us to expect groups with different parental education levels to differ by innate ability and that the level of innate ability of each group would reflect how selective education systems are. This posits that the outcome difference between two groups is simply a function of their relative

ranks within the innate ability distribution. Thus, an increase in access to higher levels of formal education should translate into decreasing innate ability among the educated.

Meritocracy and spontaneous order

On the level of society, the critical question of selection by ability for equality of opportunity can be approached through the related concepts of meritocracy and spontaneous division of labour. When substantive *equality of opportunity* prevails, all individuals have access to environments that do not put them at a relative disadvantage in the competition for social positions. The two equivalents of the ability and opportunity hypotheses – or hereditarianism and environmentalism – are known under several headings in the sociological literature. A key concept of social selection where social accidents of birth exercise no effect on outcomes, and outcomes are fully determined by ability and individual effort, is the concept of *meritocracy*. (Breen & Goldthorpe, 1999; Saunders, 1996, 2010)

This paradox of meritocracy – equal opportunities lead to new hierarchies – has been widely recognised in philosophical, sociological, genetics and education literature. Temkin makes mention of the fact that 'some believe that the equal opportunity merit principle replaces the old discredited hierarchies with a new hierarchy based on temperament and ability' (Temkin, 2013) Behavioural geneticists widely believe that with improved opportunities across society, 'genetic influences are maximized, such that educational attainment is increasingly a function of individual characteristics and less a product of social conditions.' (Ayorech, Krapohl, Plomin, & von Stumm, 2017) Michael Young (2017) emphasised the tension between the principle of full equality of opportunity and the outcome that leads all to occupy positions that fully reflect their innate abilities, giving rise to castes of innate ability. This state of affairs arises if cognitive outcomes are affected by inherited factors (i.e. innate ability) and if selection into social positions is based on cognitive outcomes.

Durkheim (1893) referred to a division of labour produced by selection mechanisms that are unaffected by social inequalities as a spontaneous division of labour. An entirely meritocratic selection produces a *spontaneous division of labour* in which social inequalities match natural inequalities. It is spontaneous by virtue of being characterised by the absence of the thumb of social or other inequities from the golden scales of selection to social positions, and thus outcomes reflect only individual abilities and not their opportunities.

> Mais par spontanéité, il faut entendre l'absence, non pas simplement de toute violence expresse et formelle, mais de tout ce qui peut entraver, même indirectement, le libre déploiement de la force sociale que chacun porte en soi. Elle suppose, non seulement que les individus ne sont pas relégués par la force dans des fonctions déterminées, mais encore qu'aucun obstacle, de nature quelconque, ne les empêche d'occuper dans les cadres sociaux la place qui est en rapport avec leurs facultés. En un mot, le travail ne se divise spontanément que si la société est constituée de manière à ce que les inégalités sociales expriment exactement les inégalités naturelles.

Durkheim recognised that no society had achieved a spontaneous division of labour but took the progressive decline of castes and towards a spontaneous division of labour to be a historical law. Durkheimian spontaneous order thus describes a division of labour arising from full substantive equality of opportunity where the only accident of birth that exercises an influence on the outcome is that of innate ability.

The Durkheimian expectation of the progress of history from unjust to just hierarchies matches what Marks (2013) has dubbed the *modernisation perspective*, which encompasses approaches that emphasise resources as the dominating factor in determining outcomes and anticipate decreased inequality when access to nutrition, education and a number of other environmental factors conducive to cognitive development becomes less unequal. From the modernisation perspective, success is mainly due to access to external resources, and the equalisation of such luck with progressive modernisation will equalise outcomes as well. What appears as modernisation is the realisation of the Durkheimian historical law away from caste hierarchies and towards a spontaneous division of labour. Thus, the modernisation perspective embodies the frequent juxtaposition of modern fair and meritocratic hierarchies and the old unfair ones. (Piketty, 2019) While Marks emphasises the historical aspect of the hypothesis, others focus on mechanisms, and then the emphasis is on social advantage and disadvantage, as in Saunders' (1996, 2010) label 'SAD thesis' to describe the opportunity hypothesis.

In Marksist terminology, the opposite hypothesis on the phase of history goes by the name of *reproduction theory.* It is a representation of the ability hypothesis from a historical perspective, trying to explain the evolution of intergenerational mobility with societal changes and policies to increase equality of opportunity. Marks (2013) posits that increasingly equal access to resources does not result in more equal outcomes because the distribution of outcomes was never crucially explained by external resources in the first place. In the reproduction theory perspective, social structures are not amenable to change because they do not primarily reflect the transmission of opportunity, privilege and luck across generations but because they reflect underlying distributions of innate ability. In this interpretation, the Durkheimian vision of modernisation from archaic and unjust hierarchies to modern and meritocratic ones never happened because the old hierarchies were already meritocratic.

A variant of this thesis is to argue that while old hierarchies may have been unjust and not meritocratic, opportunities have been sufficient, after massification of the education systems, for example, for individuals to rise or fall to a position concordant with their abilities. (Murray & Herrnstein, 1994)

When we go beyond the question of malleability of cognition, the ability and opportunity hypotheses can be understood as hypotheses on the phase of history that we are living in. The opportunity hypothesis sees that we have old, unfair hierarchies where the social accidents of birth still exercise an influence on individual outcomes. This belief may or may not combine with a Durkheimian belief in the progress in history, but it nonetheless entails that we would have a way to go before reaching a spontaneous division of labour where individuals' positions in society are determined by their abilities and effort rather than their background. The ability hypothesis posits that something akin to the spontaneous division of labour has already arisen, and thus any lack of intergenerational mobility reflects the castes of innate ability.

Opportunities left and right

The Youngian irony of meritocracy is that full meritocracy produces a society where social differences directly reflect differences in innate ability and where one would not expect significant social mobility to occur. The irony of the situation arises from the fact that the question of the spontaneity of the current social arrangements cuts to the core of the political divide between left and right. This

may contribute to the sensitivity of the discussion on equality, equality of opportunity and group differences.

It may be useful to emphasise the difference between taking outcome distributions as just, in the sense of reflecting moral merit, and taking them as reflecting merit of other kinds. We noted in ***Chapter 1*** the two dimensions of merit, being good and being praiseworthy. Being good in whatever it is that determines the position or the outcomes of an individual can be deduced from the outcomes. Outcomes have causes. It is tautological to state that the causes of the distribution are the reason for the distribution of outcomes. This is what cause means. When we take a process like evolution or competition in the market, the outcomes are also caused by something. If we take the process to be one of weighing individuals or genes on golden scales, we may say that the outcomes show that some individuals or genes had what it takes to succeed in the wild or in the market. Success is thus itself sufficient evidence of merit. This is, of course, trivially true because it amounts to asserting that outcomes have causes.

Moral merit answers the question of praiseworthiness and requires further assumptions as to what causes are constitutive of moral merit. Thus, in addition to concluding that a person, or another protagonist, has had what it takes to succeed in the competition, taking the outcome to reflect moral merit requires the assumption that the competition has been just, and thus the outcome is not tainted by unjust influences. Thus, observing an outcome does not – in the absence of further assumptions on just process – by itself establish the justice of the outcome or that we should not try to intervene in some of the causes to achieve a different outcome.

This is sometimes underappreciated due to a common conflation of two meanings of reason. We may think, following Dworkin's (2011) illustrative example, that 'Stalin had no reason not to murder his colleagues' because if he had had such reasons, they would not have been murdered. The fact that they were murdered arguably shows that there were no such reasons. At the same time, we may consider that there were excellent reasons for Stalin not to pursue his policy of mass murder. What we must recognise is that we are using 'reason' here to denote two different things. One is an answer to a call for a causal account of how an outcome came about, the other an answer to a call for moral justification of that outcome.

Norberto Bobbio (1996) has suggested that the dividing line between left and right is related to observed inequalities. The political right sees inequalities as legitimate and just, as being based on justified processes and criteria for allocating positions, as just rewards for merits. Different tendencies of the right might disagree on whether the appropriate merit is weighed by God or the market, but the conservative religious right and the libertarian right agree that observed success more or less corresponds to the merit or entitlement that has been weighed on the golden scales of justice. The left views inequalities more critically, not accepting the observed distribution of outcomes as a noncontaminated proxy for the distribution of legitimate merit and thus seeking causes of observed inequalities more from the direction of inappropriate thumbs being placed on the (at best) gilded scales.

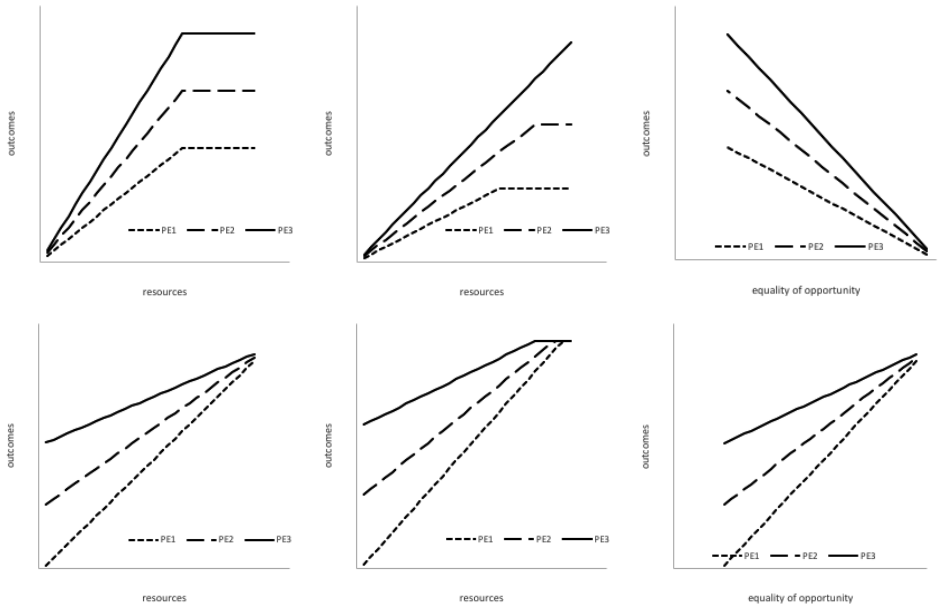
The connection of the questions of moral merit to these underlying suppositions may partially explain why the discussion on equality of opportunity is so complex, and so often riddled with accusations of ideological interpretations of the available evidence. (Gottfredson, 2000, 2004, 2013; Saunders, 1996, 2010; Sesardic, 2010) The question is located at the crossroads of two different conceptualisations of the world, and all the relevant interpretive concepts have correspondingly slightly different contents on different sides of the debate. It is to avoid this type of misunderstanding due to

insufficient elaboration of key concepts that the theoretical sections of this work are as extended as they are.

Systemic implications of modernisation and reproduction

Modernisation perspective and reproduction theory thus express the systemic content of the ability and opportunity hypotheses through their consequences, taking them to the level of Pfeffer's (2015) question on the possible trade-off between equality of opportunity and equality of outcomes. Thus, the question arises from the responsiveness of different types of individuals to environmental change, which translates into a potential interaction on the level of aggregates between equality of opportunity and the level of outcomes. The ability hypothesis posits higher marginal returns on resources spent on individuals with higher social origins, and thus higher, even though unobserved, innate ability. The opportunity hypothesis supposes higher marginal returns on individuals with lower social origins as their innate abilities have gone untapped.

Figure 1: Environment quality, equality of opportunity and the level of outcomes according to resource and ability hypotheses



The difference between these predictions is illustrated in **Figure 1**, showing fictitious examples of the overall association between environmental quality and learning outcomes predicted by the resource and ability hypotheses.

The upper row of figures illustrates the ability hypothesis, which predicts that improved environment affects the high parental education type (PE3) more than the low parental education type (PE1), as they have more unrealised potential to spare than the low parental education type. The first two display how this leads to higher socioeconomic groups benefiting more from improved resources

because their innate ability makes them more efficient at using any inputs of the learning process and places fewer limits on their cognitive development. Thus, the socioeconomic differences in outcomes should, if anything, grow with growing resources. The third figure shows this pattern whereby the higher the level of cognitive outcomes, the lower the level of equality of opportunity.

The lower row of figures illustrates the predicted co-evolution of resources, outcomes and equality of opportunity according to the opportunity hypothesis. As the opportunity hypothesis explains between-type group differences using the environment or environmental resources, the association between environment and outcomes is steeper for the low parental education type than for the high parental education type, because improved resources benefit the resource/opportunity constrained groups more. Here, improved resources lead to narrowing of socioeconomic differences in outcomes and thus to higher levels of equality of opportunity being positively associated with the level of outcomes.

The examples illustrate the basic pattern that is produced when outcomes are plotted against equality of opportunity. The ability hypothesis predicts that equality of opportunity is associated with lower learning outcomes, while the opportunity hypothesis predicts that smaller inequalities should be associated with higher learning outcomes.

The opportunity hypothesis is consistent with a trade-off between the level and equality of *outcomes,* but not between equality of *opportunity* and the level of outcomes. The ability hypothesis is consistent with a trade-off of the level of outcomes with both equalities of outcomes and of opportunity.

For the ability hypothesis, the opportunity trade-off is essentially an outcome trade-off. If the pure ability hypothesis holds, the opportunity distribution represented by financial, social or cultural resources of parents is effectively causally immaterial because the outcome distribution follows from the ability distribution. Thus, any redistribution of opportunity would be effectively equivalent to redistributing outcomes, which would necessarily lead to levelling down.

Innate ability and selection into higher education

The underlying assumption that social groups differ in their levels of innate ability is often present in education policy discussions, even when this goes unrecognised or underappreciated. This is illustrated by the discussion on massification of higher education, where the widening of access to higher education in recent decades has at times been considered to entail a reduction in the lowering of standards of selection. The reasoning is simple: if we admit more people into higher education, we must be admitting people who would not have entered under the old system and thus widening access to education must lead to lowering the required level of ability and the average ability of entrants with it. However, this reasoning supposes that screening of entry into higher education has been efficient, i.e. the selection process has effectively selected individuals by innate ability.

This presupposition has a corollary. If the selection process has been effective, the very large socioeconomic differences in the entrance to higher education, where over 60 per cent of offspring of highly educated families enter higher education against under 10 per cent of the offspring of families with low levels of education, are due to socioeconomic differences in innate ability. (Karhunen & Uusitalo, 2017) For the 10^{th} centile of one population to be equivalent in talent to the 60^{th} centile of another, the difference in the average innate ability of the populations would be on the order of 1.5 standard deviations if innate ability is normally distributed.

The question of socioeconomic differences in access and evolution of graduate ability with widening access is actually one and the same question, as widening access to tertiary education has significantly touched the lower socioeconomic groups in many countries, as the higher social strata have already had very high participation levels in higher education. Indeed, if the socioeconomic differences in innate ability are smaller than the considerable 1.5 SD differences in higher education participation, increased and equalised access to higher education has actually increased the average innate ability of new entrants rather than decreased it.

6.3 The Ability-Opportunity Curve

Combining the hypotheses

Both the ability and opportunity hypotheses postulate selection mechanisms through which substantively equal opportunity is realised so that innate ability is revealed and realised irrespective of social background. Their difference is in the level of natural between-type inequality of non-malleable innate ability that they postulate.

We may combine the two perspectives in a curvilinear association between equality of opportunity, measured without an observable measure of innate ability, and the level of achieved outcomes. The location of the peak of the *Ability-Opportunity curve* would be determined by between-type population stratification on non-malleable innate ability. Thus, the location of the peak changes with technology that affects the malleability of innate ability (if innate ability becomes entirely malleable, it becomes an opportunity) and with changes in between-type selection for the non-malleable traits. Technology should here be understood in the broad sense, including cognitive and social technologies.

Societies would move along the curve by means of different policies. On the upward slope, increased equality of opportunity may be associated with improved proficiency as long as equalisation eliminates differences based on environmental differences or differences in opportunity. Thus, the left side of the curve slopes upward because inequality of opportunity is inefficient through wasting innate ability. At the peak, any remaining differences in cognitive achievement relative to measured background characteristics are due to non-malleable differences. If any such between-type differences remain, they show up as inequalities of opportunity in standard inequality of opportunity measures, which cannot disentangle variation of innate ability associated with background.

As the peak is achieved when the environment does not contribute to group differences, it may be postulated that it is achieved when groups have realised their full biological potential. After this point, the difference in the expected value of outcomes between types cannot be narrowed by improving the outcomes of the low ability type, so the only way to narrow the gap is to lower the outcomes of the high ability type. How this may be done might cut into the outcomes of the other types as well, but as that is not strictly necessary, it is natural to think that the peak of the curve is essentially the peak of the expected value of outcomes of the high background group. On the right side of the peak, the association between equality of opportunity and the level of outcomes turns negative, as equality is only achieved by levelling down. In short: on the right side of the peak, increasing equality

of opportunity effectively turns into increasing equality of outcomes, and any equalisation is achieved through levelling down.

Thus, while for the simple ability and opportunity hypotheses, the resource and equality of opportunity axes are mostly equivalent, the Ability-Opportunity curve is different for the high-background group. For it, resources peak at the point where peak outcomes of low parental education type, or some other comparable SES group, are achieved, as the group achieves its full innate potential. After this point, adding resources does not improve their results, so the only way to narrow the difference to the high-ability/high-background group is by suppressing the achievement level of the high-ability/high-background group.

This, of course, supposes that differences between types evolve as long as the attainment of different types evolves. The outcomes are maximised at the peak of the ability-opportunity curve if the peak is achieved by maximising beneficial environment, but without strong equalisation policies. We may also imagine that all environments were equalised on a level falling short of the outcome-maximising level. If that were to occur, all differences would be explained by innate ability, but providing additional resources to e.g. the best-off group would raise their outcomes. This case may be covered if we consider outcomes to be maximised at the current level of technology and resources.

The opportunity hypothesis is to be understood as the modernisation perspective that there is a Durkheimian historical trend away from caste inequalities and towards a spontaneous division of labour where only our innate abilities exercise any control on our outcomes that is beyond our own choices and actions. The ability hypothesis, in a strong form, interprets that we are past the peak and already in the territory where outcomes are distributed according to or more equally than innate ability, necessarily turning the opportunity trade-off into an outcome trade-off, with improved equality of opportunity leading to diminished outcomes.

The Ability-Opportunity curve strengthens the rationale for choosing range as the measure of inequality of opportunity. Because it focuses on between-type differences, which are at the focus of the ex-ante approaches, it takes a value of zero when the best-off background does not produce better outcomes than the worst-off background. Thus, it can be intuitively understood as the difference between the average/expected outcomes of different types. Further, it is intuitive that the lower bound for eliminating between-type differences in outcomes is set by how much these differences are due to between-type differences in innate ability and that on the right side of the peak the ability-opportunity curve displays the outcome trade-off.

The steepness of the curve and the environmental influence

When all types, such as parental education types, are placed on the curve, between-type differences in the expected values of outcomes are placed on the same scale as the responsiveness of outcomes to environmental changes. The minimum between-type difference is, of course, achieved in the absence of type-specific limitations in access to environments contributing to the full realisation of biological potential, at the point when all types achieve their full biological potential, and whatever else of the between-type difference in expected value of outcomes is due to between-type differences in innate ability. This links the placement of the peak of the Ability-Opportunity curve on the equality of opportunity axis, determined by the association between types and innate ability, to the relative differences between the types at peak.

The steepness of the curve to the left of the peak may be taken to correspond to the responsiveness of outcomes to environmental changes. If cognitive outcomes were completely irresponsive to changes in the environment, there would be no curve. The steeper the curve, the more that environmental change can affect the level of outcomes.

This would allow us to observe how substantial the between-type differences are, relative to the differences in outcomes of the same type with different levels of resources. The relevant question here is essentially the size of the between-type differences relative to the changes in the level of outcomes that can be achieved with environmental changes. This has both empirical and normative implications. The empirical challenges are, of course, related to the challenge of assessing the role of unobserved innate between-type differences and of environmental differences that are mostly unobserved and/or very difficult to quantify when only observational data is available.

It might be noted that when genetic luck is taken to be outside human influence, it can be considered a responsibility characteristic, and consequently group differences in expected outcomes are justified if groups display differential innate ability. But responsibility characteristics may also be considered to form a continuum in which some just causes of outcomes are considered more just than others. In such a setting, innate ability might be considered a just cause of outcomes, but less just than individual choice. If effort subject to individual choice is considered more meritorious than possession of inborn traits or characteristics, such as innate ability, a new equilibrium emerges in which it is just to compensate, up to a point, bad genetic luck with environmental advantage. When this is the case, the level of legitimate compensation depends on the normative value placed on difference of effort subject to choice and innate ability, and on how responsive outcomes are to environmental changes relative to the size of the between-type differences. If very modest environmental changes were sufficient to close the outcomes gap between high-innate ability and low-innate ability groups, a more modest preference for rewarding characteristics out of choice relative to, e.g., innate ability is sufficient to legitimate full compensation for bad genetic luck with better environmental resources. When the preference to compensate choice-dependent effort over innate ability weakens or compensation becomes more inefficient (outcomes are less responsive to environmental changes), justification of compensatory allocation of environmental resources (when the level of resources is measured as the resources necessary to achieve a given reduction of the between-type difference in expected value of outcomes) is weakened.

Analysing on the curve

The ability-opportunity curve can function as a heuristic for interpreting the empirical discussion on intergenerational transmission of educational attainment or social positions as a discussion on

1) the shape of the curve for a given society and
2) the location of the society on that curve.

Traditionally, these questions have quite often been collapsed into one question, in which societies have implicitly been supposed as broadly equivalent in terms of how far social positions are genetically determined. These questions can, however, be separated as it is not clear that subpopulations in all countries have been stratified by innate ability to the same degree; if they indeed have been so stratified. And it is unclear as to how far we might be from the peak, the level of which depends on our available technology. If technology were to allow any individual to achieve the same

learning outcomes irrespective of innate dispositions, or if innate dispositions were entirely malleable, the location of the peak would be at the extreme right. Correspondingly, if humans lived in a world with no social influence on anything, the peak of the curve would be at the extreme left.

The relevant question, for any society, is where it is placed on the curve. Whether society can expect to improve both equality and efficiency simultaneously is a fundamental question of education policy. If both can be improved simultaneously, groups do not stand to lose in absolute terms if the position of others improves. It is good to note, however, that this does not mean that between-group trade-offs would completely disappear. Even if both the level of outcomes and equality of opportunity can be improved, the best-off would face a trade-off between two relative comparisons of the level of learning outcomes. Their relative advantage within a society is, by definition, narrowed if equality of opportunity improves, even if they improve in absolute terms or relative to comparable groups in other countries. How this trade-off is then dealt with depends on whether it is better to reign in hell or serve in heaven.

Associated hypotheses

The ability-opportunity curve is related to other conceptualisations of hypotheses on the contribution of innate ability on outcomes. This relationship and structural analogy should be noted, but the different conceptualisations should not be confused with each other.

First, we may note that the ability-opportunity curve – as a combination of the ability and opportunity hypotheses – is structurally analogous to the relationship of the Pareto, Saunders and Scarr-Rowe hypotheses on the heritability of – or genetic contribution to – status or cognitive achievement. The Scarr-Rowe hypothesis predicts that the contribution of genetics increases when environment improves, and thus achievement should be more heritable in higher socio-economic groups with more favourable circumstances. (Scarr-Salapatek, 1971) The Saunders (2010) hypothesis is the complete opposite, predicting that innate ability plays a greater role in the prospects of those in more unfavourable circumstances, because under better circumstances the resources available compensate for lack of innate ability by supporting progression in the schooling system and through other means that protect individuals from sliding down in the social ladder. Both hypotheses are present in the Pareto (1971) hypothesis that in low enough social strata poor conditions prevent rise despite innate ability, while in high enough strata social advantage and available resources prevent downward mobility despite lack of innate ability. Thus, the strongest effect of genes on outcomes would be felt in the middle strata, where environment allows success but does not rule out failure. If we understand the Pareto hypothesis loosely, being without commitment to the location of the peak of the curvilinear association, it becomes an umbrella hypothesis that subsumes both Saunders and Scarr-Rowe.

These three hypotheses are also closely substantively linked to the ability-opportunity curve. Scarr-Rowe hypothesis is linked to the location of the ability-opportunity curve, predicting that where it peaks, variation in cognitive outcomes is fully due to innate ability. The Saunders hypothesis is more complicated. It was originally presented in the context of an argument that, while individuals from poorer environments have equal opportunities for upward mobility in the UK, the upper strata are still protected from downward mobility by environmental factors. From the perspective of the ability-opportunity curve, it is very much like Scarr-Rowe, as equality of opportunity in both is, if correctly identified, expected to be associated with lower learning outcomes. The difference between

Scarr-Rowe and Saunders is, from this perspective, one of emphasis, as the former emphasises environmental limitations at the lower end of the socioeconomic scale, while the other emphasises environmental compensation at the upper end.

Both the curvilinear association of the Ability-Opportunity Curve and the Pareto hypothesis are examples of the explaining away effect (Pearl, 2009), where either lack of resources/opportunity or of innate ability are sufficient to prevent high outcomes, and/or the determination of outcomes by genetic differences. In both cases, the extreme locations of the peak of the curve are characterised by hypotheses where only one of the two possible causes – innate ability and environment – determines the outcome, and the location of the peak is determined by the relative importance of the two causes.

Additionally, we may note that the ability-opportunity curve framework is also linked to analyses on inequality of outcomes. Many large-scale surveys report the association of variance and the level of scores, and the decomposition of variance components into environmental and biological/innate causes of variation is common in the analysis of heritability of cognition or education. The ability and opportunity hypotheses produce similarly divergent predictions on the association between overall inequality or variance in learning outcomes and their average level. Similarly, as with socioeconomic equality, here referred to as equality of opportunity and sometimes referred to as impartiality, the opportunity hypothesis would predict that equality would be positively associated with level of learning outcomes while the ability hypothesis would predict a negative association. Any overall variation in learning outcomes follows from variation in innate ability and variation in circumstances. The opportunity hypothesis would predict that when overall variation in outcomes diminishes, the average level of learning outcomes rises, because of gradual elimination of the initial variation arising out of variable circumstances. In effect, the crucial question is the role of environmental influence in relation to biological influence. In the opportunity hypothesis, we would expect genetic and environmental advantage to be positively associated, and thus the overall variance in outcomes is greater than the variance in innate ability. The opposite hypothesis is a strong form of the ability hypothesis, assuming that environmental variation has been negatively associated with innate ability, which predicts rising outcomes when less effort is expended on compensatory measures.

6.4 Trade-offs, time and causality

Trade-off trade-off

There is also an inherent tension between the trade-offs because outcomes turn into opportunities. The outcome trade-off has been almost universally recognised, and while opinions may differ on the opportunity trade-off, the standard view has been that equality of opportunity is positively rather than negatively associated with the level of outcomes. But because the outcomes of one generation contribute to the distribution of luck of the next, the picture becomes much more difficult to untangle. The standard Okunian hypotheses of outcome and opportunity trade-offs – equality of outcomes harms incentives and inequality of opportunity wastes ability – seem very safe. The question is how to find the equilibrium at which welfare losses associated with either trade-off are minimised.

Recently, it has been shown that even outcome trade-off does not necessarily apply in the area of income distribution and economic growth. Evidence suggests that the trade-off between equality

of income distribution and economic growth disappears over the longer term , that equality of income distribution seems to be positively associated with economic growth, and that growth spells are longer when income distributions are more equal. (Berg & Ostry, 2011; Berg, Ostry, & Zettelmeyer, 2012; Ostry, 2017; Ostry, Berg, & Tsangarides, 2014)

Evidence is also emerging that the inconclusiveness of the literature on the association between income inequality and economic growth might be due to the ambiguity resulting from the fact that the distribution of outcomes is a composite of the distributions of opportunities and effort. When the inequalities of opportunity and effort are decomposed, economic growth would seem to be, as is to be expected, positively associated with equality of opportunity and negatively with equality of effort. (Marrero & Rodríguez, 2013)

The fact that levels of inequality might have an effect on the extent of intergenerational mobility has been acknowledged for a long time but has become more visible with Corak's (2013) introduction of the Great Gatsby curve, i.e. the observation that more equal distribution of income is associated with greater intergenerational mobility. Evidence is still accumulating, but thus far the finding is that the Great Gatsby curve seems to hold on the level of nations (Corak, 2013) as well as between regions within countries, at least in the United States (Chetty, Hendren, Kline, & Saez, 2014; Chetty, Hendren, Kline, Saez, & Turner, 2014) and in Finland (Suoniemi, 2017). The challenge of arguing from cross-country or cross-region correlations is evident, and China seems to display the opposite association, with regions with higher levels of income inequality associated with higher intergenerational mobility. (Fan, Yi, & Zhang, 2015)

The significance of the Great Gatsby Curve is that it links equalities of outcomes and opportunity. As the outcomes of a generation turn into opportunities for the next, higher inequality of outcomes would seem to harm intergenerational mobility, and thus harm the mobilisation of the innate ability of lower socioeconomic groups, leading to efficiency losses that Okun (2015) associated with inequality of opportunity. At least one mediating mechanism between inequality of outcomes and the opportunities of the next generation would seem to be education. A higher level of income inequality would seem to be associated with lower investment in human capital by lower socioeconomic strata, which in turn plausibly leads to lower accumulation of human capital, lower productivity and lower growth. (Cingano, 2014)

Quality-equality trade-off in education

In the field of education, different equality trade-offs have been explored using international student assessments and different adult surveys as well as various sources of data with formal educational outcomes. The increased availability of large-scale survey data on observed learning outcomes has contributed to the emergence of international literature on the relationship between equality and learning outcomes.

The research using student assessments has overwhelmingly used data from the Programme for International Student Assessment (PISA), Trends in International Mathematics and Science Study (TIMMS) and to a lesser extent the Progress in International Reading Literacy Study (PIRLS). Research using data on adult skills has used both skills surveys that directly assess learning outcomes, International Adult Literacy Survey (IALS) and Programme for the International Assessment of Adult Competencies (PIAAC), but also social and labour market surveys that collect data on formal educational attainment.

In the research using international student assessments, no clear association between the levels of equality or equality of opportunity and the level of outcomes has been established. (Werfhorst & Mijs, 2010; Woessmann, 2004, 2009) In general, the results from PISA have effectively questioned the existence of equality-efficiency trade-offs, both the outcomes trade-off and the opportunity trade-off, in the area of education.

There is evidence of weak to moderate positive correlation between equality and level of learning outcomes, but with many countries scattered along both dimensions. The OECD has become progressively more unequivocal in its assertions as to the lack of a trade-off between equality and performance. With PISA 2003, the OECD concluded that 'wide disparities in student performance are not a necessary condition for a country to attain a high level of overall performance' (OECD, 2004). The conclusions became progressively bolder with PISA 2006 and 2009 (OECD, 2007; OECD, 2010, OECD, 2013a). For PISA 2012 (OECD, 2013b) the conclusion was that 'for most countries and economies, improvements in performance need not come at the expense of equity' and one volume of the study bore the subhead 'Excellence through Equity.' All in all, the OECD interpretation of PISA has lent support to the interpretation that policy does not face an equity-efficiency trade-off in the field of education, at least compulsory education.

The overall association between the level of equality of opportunity and the level of proficiency is modest in PISA and in the other large-scale surveys, with many countries having high performance and low levels of equality or high equality and low levels of performance. (OECD, 2004, 2007, 2010, 2013) Nathkov and Kozina (2012) found that an increase of equality of educational opportunity by one standard deviation (SD) is associated with an increase in average test scores by 0.3 SD in PISA 2009, controlling for GDP per capita.

Evidence from student assessments may not be automatically generalised to the adult population. The youth of any given moment in time is not representative of the entire adult population, most of which has had, even in developed countries, very different levels of access to education than the current youth, with the older cohorts generally enjoying lower levels of access. Additionally, education systems change over time, which also would seem to affect the learning outcomes of different cohorts, and different cohorts grow up in different environments in general as well. And we know from, e.g., the Flynn effect that environmental change may have considerable effects on the level of cognitive outcomes.

There is some research on the association between equality of opportunity and the level of learning outcomes using data on adult skills, which has found a weak but positive association between equality of opportunity and learning outcomes using data on adult skills, namely the International Adult Literacy Survey (IALS) and date from the PIAAC. (Pfeffer, 2015; Willms, 1999, 2003) In the PIAAC this association is not statistically significant, and many countries show above-average literacy scores while also having steeper-than-average socio-economic gradients, while others combine below-average literacy scores with flatter-than-average socio-economic gradients. (OECD, 2013b, 2013c) The first Programme for the International Assessment of Adult Competencies (PIAAC) report finds the level of equality positively associated with the level of outcomes. 'Seven countries […] combine above-average literacy scores with a socio-economic gradient that is flatter than the average, and six countries […] show below-average literacy scores and a steeper-than-average socio-economic gradient.' (OECD 2013) However, the report shows that the link is far from conclusive, as there are many countries that do not fit the purported link. 'The Czech Republic, Finland, Flanders (Belgium) and the Slovak Republic have above-average literacy scores while also having a steeper-than-average

socio-economic gradient, while some countries, including Denmark, Ireland and Korea, combine below-average literacy scores with a flatter-than-average socio-economic gradient.' (OECD 2013c) In the report, a rise of the socioeconomic gradient by one point is associated with a decrease of 0.72 points in average proficiency ($R^2 = 0.0938$).

The shortcoming of much of the existing research has come from analysing the systemic association of equality and proficiency using adult data and taking jurisdiction as the unit of analysis, even though in most jurisdictions different age groups differ markedly in their levels of proficiency and equality, of both opportunity and outcomes. This is hardly surprising in the light of the different general environments, educational systems and levels of access enjoyed by different cohorts in different countries. Furthermore, in PIAAC, the high number of jurisdictions with high performance and steep socio-economic gradient or low performance and low gradient seems to arise from cohort differences and compositional effects. For example, in Finland, which on a jurisdiction level combines above-average results with a steeper than average socioeconomic gradient, the oldest and youngest age groups have steep socioeconomic gradients and lower performance, while the 25-54-year-old age groups have both high proficiency and very flat socioeconomic gradients. The association of flat socioeconomic gradients with high proficiency is masked on the country level by taking the whole adult population into the analysis as one case.

If skills are to a considerable degree affected by education, most of which takes place during the first two to three decades of life, and if the educational system changes over time and the level of skills changes with age, the adult population should effectively be divided into subpopulations in which their proficiency and the distribution of learning outcomes should be a function of their initial educational experiences at youth and the change/deterioration in their cognitive skills through adulthood. (Paccagnella, 2016) Aggregation of these subpopulations without controlling for systematic differences leads to problems of interpretation. This study addresses the shortcomings of the previous research by treating age groups as different populations and performing country-level analysis using aggregates of standardised age-group level data.

Moreover, practically the analyses of trade-off invariably explore the trade-off between quality and equality on the aggregate level, with no consideration of the fact that the relationship of equality of opportunity with the level of learning outcomes depends on the type or group. For types from less favourable backgrounds, level of outcomes will be positively associated with improved equality of opportunity under very much less stringent conditions than for high-background types. Thus, predicting quality in the aggregate, without distinctions by type, complicates interpretation as aggregate proficiency is also affected by compositional issues because some parental education types, for example, are more common in some countries than others.

Trade-offs and the question of causalities

A case was made above for more explicitly considering measurement of equality of opportunity a programme of causal inference, because the causality is essential for the interpretation of the results. Now that we have taken a brief look at different trade-offs, it may be useful to note that the question of causal inference in inequality of opportunity may be linked to an insufficient differentiation of different trade-offs in the literature.

The economist option of downplaying the nature of measurement of equality of opportunity as a case of ill-defined causal inference is an unstable balancing act because inequality of opportunity is

about the causes of the distribution and the results only acquire a meaningful interpretation through causality. Solutions at either extreme seem unwelcome here. It is clear that unobservable variables play havoc with the causal interpretation of measures of intergenerational mobility as measures of equality of opportunity if we take the circumstance variables to represent well-defined interventions that can act as direct handles for policy interventions. On the other hand, foregoing mention of opportunity would be an overreaction, because it would mean that opportunity would have to be excised from the language of empirical enquiry for the foreseeable future. Opportunities are always hypotheticals with no physical existence, because they are the paths that were not taken. That is why they require a causal model that can be used to construct counterfactual distributions. There is always uncertainty in the structure of a causal model, even with access to experimental data. In the use of solely observational data, it is clear that full certainty, if it can be achieved, is quite a way off. It would be reasonable, in these circumstances, to conclude that full certainty is unattainable in empirical enquiry. As Dworkin (2011) noted: 'Absolute confidence is the privilege of fools and fanatics'.

The middle road would be to use the language of opportunity and causality, because that is the object of interest and of enquiry, but to make clear the reservations involved. First, it should be emphasised that the estimates are conditional on correct identification of the causal model and the absence of confounders. Most notable among the possible confounders is, of course, innate ability, to which we turn to in the next chapters.

Second, it should be emphasised that the causes of interest are not well-defined, and thus the questions do not concern the mechanisms of intergenerational transmission. This lack of definition overlaps partially with the problem of confounding variables. Indeed, if part of the differences between parental education groups in educational outcomes were explained by between-group differences in innate ability, one part of the ill-defined cause of parental education would be innate ability associated with background. However, parental education can act as a proxy for a myriad other proximate or more ultimate causes of outcomes. Even if all of them were environmental, the ability to proxy them with parental education does not entail the ability to affect the causes of unequal opportunities by means of policy.

Summary

Here we take the Ability-Opportunity Curve as a heuristic framework that collects the competing perspectives and hypotheses of the origin and nature of between-group differences in cognitive achievement, and thus provides a framework for the potential trade-offs involved. This framework presents the hereditarian or nativist perspective of the ability hypothesis and reproduction theory on the same continuum as the environmental emphasis of the opportunity hypothesis. In this way, it emphasises the contextual nature of the question and the fact that the relative importance of innate aptitudes depends on the environment, and demonstrates how the opportunity trade-off arises from a combination of outcome trade-off and selection of social groups by innate ability.

Most importantly, it allows a simple and intuitive way of discussing the ability and opportunity hypotheses and the related trade-offs, while drawing attention to the fact that we are not, in general, only interested in the strength of association of observed socioeconomic status with unobserved innate ability, but on whether the current differences in attainment are smaller or larger than what would emerge if all socioeconomic groups were to have access to equally optimal environments. This is the

crucial question not only for policy but for the practical application of normative criteria to the current situation in any one society.

Our approach takes measurement of inequality of opportunity to be a macro-level question with limited microfoundations, i.e. uncertainty in terms of mechanisms. While we need a micro-level understanding of the mechanisms of intergenerational transmission of position, we also need research on how our imperfect measure of equality of opportunity behaves in relation to other policy objectives of interest, such as the level of outcomes, and how it changes over time or reacts to policy changes. Such work may provide us with clues as to the mechanisms of transmission, lending more credibility to some explanations than others.

Clarity in this respect would be especially crucial for policy questions such as Pfeffer's fundamental trade-off. The hazards of outcome trade-off are well known, and it is widely recognised that achieving equality by levelling down is hardly welcome. But opportunity trade-off is a very different kettle of fish and requires fundamental differences in innate ability between the relevant groups – genders, races or socioeconomic groups. Clarity regarding the underlying assumptions of policy choices would be very welcome if policy were based on such assumptions.

Next, we turn to what current evidence shows about the innate or biological origins of the outcome differences between social groups.

7 Genetics of sex, race and class

Now, where are we, in terms of understanding how observed differences between different groups have their origins in innate, non-malleable differences between the groups or in the circumstances faced by the different groups?

7.1 Groups, heritability and change

Group differences

Turkheimer (2000) has declared that the 'old-fashioned nature-nurture debate was about whether or not genes influence complex behavioural outcomes, and that question has been decisively answered in the affirmative', when he proposed three laws of behaviour genetics. This summary has stood for the last two decades (Chabris, Lee, Cesarini, Benjamin, & Laibson, 2015) and it is now widely recognised that genes contribute to more or less all behavioural traits. (Plomin & Deary, 2015)

It may be argued, however, that, despite apparent agreement, the fundamental tension of the old nature-nurture debates has not been resolved. Turkheimer's laws have been widely accepted, but there is a vast range of positions in academic discussion on the role of genetics in explaining outcomes on the individual and especially group levels. For the most part, the continuing disagreement in the field of cognitive genetics does not seem to be about empirical findings, but about their interpretation. Arguably, this is because much of the old nature-nurture debate was never about individuals, but about groups. It has been less controversial than sometimes presented that individuals possess innate differences in physical and mental characteristics or dispositions. Sensitivities relate to groups. We are very often interested in the role of innate characteristics and environment in producing the outcomes of different groups that differ not only by their environment but – more or less, depending on the groups in question – genetically as well. Indeed, it is through our interest in between-group differences that the matter is relevant for the work at hand.

The relationship between environmental and genetic differences may depend on the type of groups in question. Possible genetic sex differences in level, distribution or composition of cognitive abilities are qualitatively different from some other group differences as we do know that the sexes are genetically different, with parts of the genome having a very long separate evolutionary history. In addition, the sexes display clear, if not usually very pronounced, differences in the distribution of different cognitive or behavioural traits. (Halpern, 2000) The separateness of the evolutionary histories of other populations depends significantly on the populations in question. Some populations have been relatively isolated for as much as thousands or tens of thousands of years, which gives ample time for genetic differences between that population and others to emerge. Some of these populations have been geographically isolated, some only isolated as populations despite regular contact. Relatively isolated populations, without geographic isolation, can be found not only in Africa but also in India, where castes have also been genetically relatively isolated. (Reich, 2018) A case could be made that geographically distant populations have different evolutionary histories, and their genetic differences might, therefore, contribute to cognitive differences between the groups.

On the other hand, we have socioeconomic groups (however socioeconomic status is measured) that have may have had a much shorter existence as distinctively separate populations, and for which the very question of the separateness of the population is very much more uncertain. Where the differences in the biological characteristics of such populations cannot have slowly emerged over evolutionary time, they must have emerged through selection mechanisms by which individuals would 'would quickly rise above his family or sink below it, and take his place, both biologically and socially, with his peers.' (Herrnstein, 1971)

Sensitivities in the area of group differences are rather unsurprising, as the groups with supposedly lower innate ability are those that have historically and contemporarily suffered from more impoverished environments, even downright oppression and victimisation. Justification of gender differences in attainment by innate ability rings hollow when women have even been formally excluded from full citizenship, ownership rights and educational opportunities. Justification of racial differences in attainment by innate ability rings hollow when some racial groups have suffered from long histories of slavery, oppression and deprivation. Justification of socio-economic differences by innate ability rings hollow when some social groups have suffered from much material deprivation and lack of access to education. The impression of hollowness has not been helped by the fact that the inferiority of some groups has been a more constant feature of the analyses than the grounds of that inferiority. Grounds have shifted, but the outcome has not. (Kevles, 1995; Rippon, 2019)

Unfortunately, when it comes to group differences, we are far from having conclusive evidence of the innateness or malleability of the causes of cognitive outcomes. For our purposes, the critical question is the causal role of genetics in explaining the variation of outcomes between groups in different circumstances. Standard treatments of inequality of opportunity take group membership to be a treatment. As noted, it is an ill-defined treatment, and for both practical and normative perspectives it would be important to know to what degree the treatment in question is genetic and to what degree it is environmental.

The following treatment does not attempt to resolve the issue of the innateness of group differences in ability, but it addresses some of the relevant questions to put our empirical results in perspective.

Heritability of cognition - all behaviour is heritable

We may start by noting that there is a widespread consensus that all behaviour is heritable. The classical methods of population genetics allowed the identification of the heritability of traits without genetic data, from phenotypic data. This pedigree-based era of twin and adoption studies very effectively statistically decomposed the variation of outcomes based on known differences in genetic and environmental similarity between different types of twins. The higher similarity of identical/monozygotic twins with each other than of fraternal/dizygotic twins has been taken as an indication of the contribution of genetics. Identical twins share all of their genes, while fraternal twins share half of their genes on average. Thus, as demonstrated in ***Equation 0,*** the difference in similarity – correlation of outcomes – should be the contribution of half of the genome and twice the difference would be the contribution of genetics to the overall variation of outcomes.[17]

[17] Of course, it is an assumption that the environment of identical twins is not more similar than the environment of non-identical twins. If it is, as has been argued (Lewontin, 1996), the decomposition does not work. Indeed, we know the assumption to be false, as identical twins enjoy more similar environments than non-identical ones even in the womb. The

$$h^2 = 2(r_{MZ} - r_{DZ}) \qquad (0)$$

What emerged was an understanding – enshrined in Turkheimer's (2000) first law of behaviour genetics – that 'all human behavioural traits are heritable. 'Indeed, meta-analysis of practically all twin studies of complex traits published 1958-2012, covering some 17 000 traits in 1748 studies and with over 14.5 million partly dependent twin pairs, did not find a single trait with 'weighted heritability estimate of zero' and found the traits to be 49 % heritable on average. (Polderman et al., 2015)

The twin studies era provided us with reasonable estimates of the heritability of cognitive traits by exploiting the differences between mono- and heterozygotic twins reared together and apart, as well as the degrees of similarity of children with their biological and adoptive parents. Hundreds of studies over decades provided a longstanding body of evidence showing that cognitive traits are highly heritable (Eysenck, 1979; Haier, 2016) and that intelligence is to a considerable degree, from 0.4 to 0.7, genetically inherited, and the heritability of intelligence rises with age. (Haier, 2016) Thus, intelligence would seem to be less heritable than characteristics such as height, but quite heritable nonetheless, with heritability estimates quite close to the average of behavioural traits.

With the rapid development of molecular behavioural genetics, there is hope that we might be able to progress from phenotypes and pedigrees to come closer to understanding how genes affect behavioural outcomes. The newer phase of molecular behavioural genetics started with an early phase of studies trying to identify genes that affect cognitive or behavioural outcomes. The search was for candidate genes that would have exercised relatively large effects on behavioural traits, and the studies came up empty because such genes with relatively large effects were not there to be identified. The search for associations led to increasing sample sizes that would allow detection of ever smaller statistical effects, giving birth to genome-wide association studies (GWAS), with extensive international cooperation allowing the collection of huge samples that allow the identification of previously unobservable small effects.

The era of molecular behavioural genetics has not changed the consensus that all traits are heritable, but now we can find molecular correlates for behavioural differences. What has been revealed is that behavioural traits and other complex traits are radically multigenic, with very small effects for individual genes. (Plomin & von Stumm, 2018) Indeed, the so-called omnigenic model of inheritance has even been proposed. (Boyle, Li, & Pritchard, 2017)

The radical polygenicity has given rise to polygenic scores (PGS), calculated from the observed associations of single nucleotide polymorphisms (SNPs) with the outcome of interest, and aggregating the small statistical effects of a very large number of SNPs to form a composite score that can be used to statistically predict the trait of interest.

nature of the assumption is often glossed over in more popular – or sometimes not even very popular – accounts that emphasise how the difference between MZ and DZ twins 'The only possible reason for this is genetic: after all, the only thing that differs between the two types of twins that would make them more similar – so long as each pair is raised in the same family – is the percentage of genes they share.' (Ritchie, 2015) Happily, the era of molecular behavioural genetics is here to help, as it allows for measurement of heritability directly from molecular data and is thus free from the underlying assumption of the treatment of identical twins. Thus, we may eventually see to what degree the assumptions of twin studies led to overestimation of the heritability of cognition.

Genetics of education and social position

The genetics of intelligence suffers from the fact that intelligence has to be psychometrically measured, which limits the size of samples with both comprehensive genetic and intelligence data. Plomin and von Stumm (2018) thus see that the breakthrough of cognitive genomics appeared through molecular genetic prediction of educational attainment. Educational attainment has been one of the standard questions in many surveys, and thus self-reported data on attainment[18] has been available for a much larger population than intelligence data. Overall, some 40 % of the overall variation in educational attainment has been estimated to be explained by genetics. (Rietveld et al., 2013) For analysis with SNPs, the contribution of genetics to educational attainment has been estimated to be at least 20 % (Okbay et al., 2016).

Recent work also allows assessment of genetic influence on intergenerational continuity of socioeconomic status (SES) or educational attainment. Marioni et al. (2014) have estimated that 21 % of the variance of educational attainment and 18 % of socioeconomic status is explained by common SNPs. Further, 26 % of the variation of SES explained by genetic variation was shared with genes that influence intelligence. They further find that

> In conclusion, using DNA SNP data we found modest genetic contributions to differences in socioeconomic status that have measurable association with the genetic basis of intelligence differences and education. The implication from these analyses is that socioeconomic status and education have, in part, genetic causes that are shared with the genetic contributions to measured intelligence.

Similarly, it has been reported that while SNPs explain 2.5 % of the variation of family SES and 3.0 % of the variation of educational attainment, 'approximately half of the phenotypic correlation between children's family SES and their educational achievement was mediated genetically.' (Ayorech, Krapohl, Plomin, & von Stumm, 2017; Krapohl & Plomin, 2016)

Research on the genetics of social status and class are also increasingly interested in income as a variable to be genetically explained. Recent evidence from GWAS with UK Biobank data suggests that genetics explains some 7 % of income differences (Hill et al., 2019), while twin data from Finland would put the genetic contribution to intergenerational heritability of earnings at some 40 % for women and half for men. (Hyytinen, Ilmakunnas, Johansson, & Toivanen, 2019)

Malleability of cognition

But heritability is not the same as non-malleability. A natural starting place for a brief look at malleability of cognition is intelligence, as there has been some inherent tension in the field of intelligence research since the very beginning, with some emphasising intelligence as capabilities that can be acquired while others emphasised intelligence as a more fixed trait or tendency. (Eysenck, 1973, 1979, 1998) Indeed, intelligence tests have been developed to assess the ability to learn, reason and solve problems, to serve selection and diagnostic purposes in educational institutions and the military. At the same time, they have tried to assess a permanent individual trait or tendency, and intelligence

[18] Educational attainment is standardly measured as years in education, but it is a measure of the highest qualification attained, rather than a measure of the time spent in an education institution. Years in education are a way of presenting qualifications data on a continuous scale of educational attainment. Moreover, the existing qualifications typologies are to a considerable degree based on the typical duration of studies required for a given degree.

is very often conceptualised as a permanent, non-malleable trait. It may be of interest to the present discussion that many of the proponents of the ability hypothesis have relied heavily on intelligence research in making their case for the dominant role of innate characteristics in the determination of group differences in cognition.

But intelligence research has also very effectively called into question the view that intelligence is non-malleable, innate and fixed. Famously, this occurred on a large scale when evidence emerged of a long secular rise in measured intelligence. The prolonged rise of IQ scores has been observed in many countries in the 1900s and 2000s. The rise was baptised the Flynn effect in acknowledgement of the role of James Flynn (1984, 1987) in bringing the effect to general awareness.[19] The Flynn effect has been observed in over 30 countries, ranging from the United States and other developed countries to many developing countries (Flynn, 2012).

Many aspects of the rise have been controversial, Researchers emphasising intelligence as a biological, genotypic limit have tended to interpret that 'it is not intelligence as such that is increasing, but merely a not perfectly accurate measure of it' (Eysenck, 1979), and those who emphasise the unity of cognitive capacities (very often the same researchers) interpret the improvement as being in specific skills rather than in general intelligence *g*.

The long-term rise in IQ scores has been on the order of 0.3 points per year, corresponding to an improvement of 3 points per decade and one standard deviation per half-century. Over a century, this pace corresponds to a given level of performance reached by some 15 % of the population (1 standard deviation above average) at the beginning of the century being reached by almost 85 % (1 standard deviation below average) at the end of it. The speed of the rise in scores is thus non-negligible and the relative standing secured with a given level of performance changes significantly over time. The speed of the rise in performance in intelligence tests would seem to rule out genetic explanations for the Flynn effect, and there is a broad consensus that the effect is environmental in origin. (Flynn, 2012; Pietschnig & Voracek, 2015; Rindermann, Becker, & Coyle, 2017; te Nijenhuis & van der Flier, 2013; Trahan, Stuebing, Fletcher, & Hiscock, 2014) There is still uncertainty as to the exact environmental reasons behind the rise. The factors considered range from general environmental change, nutrition and medical improvements to education, but no consensus seems to be emerging as yet.

For some countries, the rising IQ scores of the Flynn effect have turned into the declining IQ scores of the anti-Flynn effect. These countries include almost all of the Nordic countries and the Netherlands, United Kingdom and Australia. (Dutton & Lynn, 2013; Shayer & Ginsburg, 2009; Shayer, Ginsburg, & Coe, 2007; Teasdale & Owen, 2008) For the anti-Flynn effect, there has been more uncertainty as to the reason of observed decline in intelligence.

The anti-Flynn effect has sometimes been explained using dysgenic fertility, a hypothesis that was less prominent in intelligence discussions for a while. Dysgenic fertility is the hypothesis that intelligence is declining or will decline due to higher-fertility low-intelligence individuals, most often taken to be members of lower social groups, and has been present in eugenics and education discussions over the 20th century. (Kevles, 1995) Unsurprisingly, dysgenic fertility has been raised as a possible explanation of the anti-Flynn effect. (Rindermann et al., 2017) The rise and fall of intelligence can, in this interpretation, be due to two competing forces: genetic and environmental. On the

[19] Murray and Herrnstein also note that the phenomenon that was first identified in the 1930s and the first research evidence on the effect was published in the late 1940s and early 1950s. (Cattell, 1950; Maxwell, 1954; Tuddenham, 1948; Murray & Herrnstein, 1994).

genetic side, differential fertility, with low-intelligence individuals having higher fertility than high-intelligence individuals, has exerted a constant negative pressure on genotypic intelligence. (Woodley, 2012) On the other hand, environmental factors have been driving the Flynn effect. The turning of one effect into another has been argued to arise from the gains brought by environmental changes petering out and the effects of the previously hidden intellectually dysgenic trend consequently becoming visible. (Rindermann & Becker, 2018; Rindermann & Thompson, 2013; Woodley & Meisenberg, 2013) In this view:

> As this is ultimately underpinned by intelligence, it would have a genotypic limit and if genotypic intelligence were declining then the imperfect nature of the IQ test as a measure of intelligence would mask this, but only up to the genotypic limit. Once this limit was reached, any genotypic decline in IQ would become visible on the IQ tests. (Dutton, van der Linden, & Lynn, 2016)

Recent evidence with Norwegian conscript data for the 1962-1991 cohorts is inconsistent with the dysgenic fertility hypothesis, as 'the observed Flynn effect, its turning point, and subsequent decline can all be fully recovered from within-family variation.' They find 'no evidence for prominent causal hypotheses of the decline implicating genes and environmental factors that vary between, but not within, families.' (Bratsberg & Rogeberg, 2018)

The Flynn effect has conclusively shown that environmental change may cause a change in the level of intelligence and that the magnitude of the change is very notable relative to the distribution of intelligence at any given point in time.

The malleability of intelligence by intervention

Why has intelligence changed so much, and can it be changed by purposeful intervention?

While innate ability is by definition non-malleable, observed intelligence has been shown to be malleable and to be influenced by such environmental factors as nutrition, adoption and education. While a full review of the literature is out of reach here, we may note some important lines of evidence.

First, adoption into a better environment seems to raise the intelligence of adoptees, on average. It is very consistently found that offspring intelligence is more strongly associated with the intelligence of their biological parents than of their adoptive parents, which would indicate that biological factors are at work in determining intelligence. However, adoption has also been associated with non-negligible gains in intelligence. The observed effects range from very high to quite modest. One SD raises have been reported by Schiff et al. (1978) for a small sample in France. Much more modest but non-insignificant gains of 4.4 points have been reported with an extensive Swedish data set. (Kendler, Turkheimer, Ohlsson, Sundquist, & Sundquist, 2015)

Adoption is of course, on the level of a population, a rather rare form of intervention to raise cognitive capacity. A much more common IQ-raising intervention would seem to be education. Ceci (1991) provided an extensive review of mostly observational literature on the effects of education on intelligence and suggested that each year of additional education contributes some 1.9 points to intelligence, an estimate which was remarkably close to a much earlier estimate by Härnqvist (1968) of an increase of 1.8 points in intelligence tests per one year of education. (Gustafsson, 2001)

The observational studies on the effects of education obviously struggled with the challenges of unobserved variables. Fortunately, a number of studies using quasi-experimental designs have been able to assess the causal effects of discontinuous changes in education, i.e. education policy reforms, on intelligence. The studies achieve quasi-experimentality by three different means: 1) controlling for pre-intervention intelligence; 2) using compulsory schooling law as an instrumental variable; or 3) using regression-discontinuity design around the age of school entry. The results indicate that, with methodologically robust identification that should reveal causality, one year of additional education would seem to lead to a 1-5 IQ point increase that also persists with age. The results imply that observational studies did give wild overestimates of the effects of education.

As the standard IQ measurement scales set one standard deviation at 15 points, 1-5 points per year of education are not insignificant, especially when younger cohorts in developed countries currently receive some 15 years of formal education on average. The authors note that 'education appears to be the most consistent, robust, and durable method yet to be identified for raising intelligence. (Ritchie & Tucker-Drob, 2017) [20] Given the prevalence of education, formal education is most likely the most effective politically controlled and effective intervention that is already implemented on a large scale with the express purpose of raising the cognitive performance of the population.

There are numerous other interventions, from the very specific to the somewhat more general. One example of the robust effects of environment on cognitive ability is learning to read, which seems to improve cognitive performance not only in the verbal domain but also in general intelligence. (Ritchie, Bates, & Plomin, 2015) Further, as Ritchie and Bates (2013) note, childhood reading and mathematics ability explain variation of socioeconomic status in adulthood beyond intelligence or socioeconomic status at birth.

Obviously, observed intelligence cannot be taken to be a measure of genotypic or innate ability and thus used to deconfound the effects of the environment on outcomes, as it is also confounded by environmental causes. The literature on inequality of opportunity has rarely used genetic data, but it has been interpreted (Ramos & van de Gaer, 2016) that Björklund, Jäntti & Roemer (2012) use IQ as a measure of 'genetic luck'. This is something of an overinterpretation, as the article in question refers to IQ 'partially or substantially' reflecting inborn traits but does not present observed intelligence as a measure of non-malleable innate ability.

[20] For example, Ceci has raised the hypothesis that rising IQ would be due to rising levels of schooling of the population over the 20th century. The estimates are very much approximate, but if one year of education adds 3 points to IQ, then an additional five years of average years of schooling would raise IQ by one standard deviation. The number of years of schooling in advanced countries increased somewhat over 3 years between 1900 and 1950, and slightly under five years between 1930 and 1990. (Barro & Lee, 2015) Thus, the realised increases in the years of schooling and in IQ scores are consistent with the quasi-experimentally estimated causal effects of education on IQ. However, it is very much necessary to note that there is no consensus on the causes of the Flynn effect. Thus, while the results of Baker et al. (2015) suggest that formal education may have been a significant contributor to the Flynn effect, this author is not suggesting that the match between the three pieces of data should be interpreted as evidence that the Flynn effect is produced by formal education. It has also been suggested that the hypothesis of education as a possible contributor to the Flynn effect is further supported by the fact that gains in fluid intelligence seem to be greater than gains in crystallised intelligence, while education has been associated with larger increases in fluid than crystallised intelligence. (Kanaya & Ceci, 2011)

7.2 What is, what could be

Weak and strong biological explanation

The big question, however, is: so what? How does the persistent heritability of all behavioural traits or apparent malleability of intelligence help us find the relevant causal counterfactuals when it comes to group differences in behavioural outcomes? Well, most of the time, without additional assumptions, it does not. We are still very much dependent on the triangulation of different types of circumstantial evidence to make reasonable assumptions on any genetic, non-malleable basis for group differences in cognitive performance.

While much of behavioural genetics has been devoted to assessing the heritabilities of different traits, it is common practice to emphasise, like Plomin (2018), that heritability 'describes what is but does not predict what could be.' It is a well-established tradition that writers lament the common misunderstandings of the concept of heredity. It does not describe the malleability of a trait, and traits with very high heritabilities, such as height, have shown huge changes due to changing environments (Hauspie, Vercauteren, & Susanne, 1997), as has also been well recognised in the hereditarian literature. (Gottfredson, 2011) Heritability does not measure the contribution of genetics to the outcomes of an individual, but the association between variation of phenotypic outcomes and variation of genetic differences. As Plomin (2018) notes, it is 'a statistic that describes a particular population at a particular time with that population's particular mix of genetic and environmental influences'.

Because heritability is environment- and population-specific, it includes the interplay between genetics and environment, which poses considerable challenges for the interpretation of the heritability coefficients. While their interpretation is not as easy as often believed even for animals whose environment can be controlled, it is very challenging for free-ranging humans in environments that cannot be controlled and are even observable or quantifiable only to a very limited degree. (Turkheimer, 2012; Wray, Kemper, Hayes, Goddard, & Visscher, 2019)

Heritability does not tell us what can be, due to what Turkheimer (1998, 2016) calls the difference between weak and strong biological explanations. Turkheimer (1998) defines weak biologism as a philosophical position arising directly from materialism and referring to the recognition that all social and psychological phenomena are characteristics of humans as biological beings. Because humans are material and biological beings, it would be surprising if our cognitive and behavioural traits would not have correlates in our material and biological substrate. A weak genetic explanation is thus constitutive of

> the observation that, one way or another, genetic differences among people wind up correlated with phenotypic differences. (Turkheimer, 2016)

Heritability, as a concept, is thus situated in the field of weak biological explanation. It establishes that the observed differences in behaviour or other traits are associated with biological, genetic differences between individuals. Turkheimer (1998) contrasts this weak version with strong biologism, in which biological

> denotes behaviour that is well-represented as a process at a biological level of explanation. An empirical discovery that a high-level human behaviour is biological entails identification of a structurally or functionally localised biological process that explains some large part of the high level phenomenon.

In the strong genetic explanation, 'observed phenotypic difference is a manifestation of a specific latent genetic mechanism.' The strong genetic explanation is not synonymous with naïve genetic reductionism, where an outcome would always be determined by a gene or a set of genes, but that strong genetic explanation may also be probabilistically deterministic, as genes are expressed in the context of a genome and in interaction with the environment.

Empirically, the whole history of behavioural genetics 'can be seen as an extended attempt to proceed from weak to strong genetic explanation'. (Turkheimer, 2016) This explains the high hopes of the 1990s of going beyond heritability, to better understand the biological and social basis of the heredity of behavioural traits. In the language of our preceding chapters, it is a question of progressing from ill-defined causality – where a causal relationship is assumed, but the mechanism is unclear and complex – to well-defined causality.

GWAS

Where the era of very large-scale studies of molecular genetics excels is the identification of associations between genetics and different phenotypic outcomes. The large sample size allows detection of minimal statistical effects, which can then be used to construct polygenic scores that predict variation in the trait of interest. The twin-era was able to show that all behavioural traits are significantly heritable. (Polderman et al., 2015) To say all is no exaggeration.

> Not only the major and established dimensions of behaviour turned out to be heritable, but so did everything else. Depression is heritable, but so is marital status; intelligence is heritable, but so is how much TV people watch. (Turkheimer, 2012)

The challenging question for research is: what do all these associations mean? Before the era of mass molecular genetics, Lewontin (1996) warned against misinterpreting heritability by noting that

> the two social traits that have the highest resemblance between parents and children in North America are religious sect and political party. Yet even the most ardent biological determinist would not seriously argue that there is a gene for Episcopalianism or voting Social Credit.

The GWAS era has allowed the discovery of significant associations from the molecular level to behaviour for very much all behavioural traits studied. Instead of a result that everything is heritable, we can have specific genes associated with Episcopalianism, watching TV and voting for a particular party. The challenge is understanding, in biological terms, what these associations signify.

Due to huge samples, now often surpassing the million mark, they are capable of reliably detecting very weak associations. What can be celebrated or lamented is the fact that polygenic scores have predictive power even in the absence of knowledge of the intervening mechanisms by which the genetics are turned into observed outcomes. Plomin (2018) is very upbeat about the positive side, i.e. the predictive ability of the polygenic scores:

> It is remarkable that polygenic scores can predict psychological traits without knowing anything about these intervening processes.

But Turkheimer (2012) is decidedly less enthusiastic, noting that by achieving predictive power without understanding the intervening processes, polygenic scores are achieving this on the cheap, by

relinquishing the quest for a strong biological explanation. Here, his difference of opinion with Plomin is not so much on the observed state of the world, but on the interpretation of the implications of those observations. After all, Plomin (2018) is aware of the shortcomings of the PGS when it comes to understanding mechanisms of inheritance.

> However, the correlation between a polygenic score and a psychological trait does not tell us about the brain, behavioural or environmental pathways by which the polygenic score affects the trait. A long slog up these pathways will be required to understand the intervening processes, especially because tens of thousands of DNA differences are involved, each with very small and highly pleiotropic effects.

Genetics in a changing environment

As noted, heritability is an environment- and population-specific measure. When the environment differs more in the availability of food, for example, the genetic influences on height explain less of the height differences between individuals than when everyone has access to adequate nutrition.

Behavioural geneticists widely believe that with improved opportunities across society, 'genetic influences are maximized, such that educational attainment is increasingly a function of individual characteristics and less a product of social conditions' (Ayorech et al., 2017) and it has been suggested that heredity could thus be interpreted as a measure of equality of opportunity. (Asbury & Plomin, 2013; Tucker-Drob & Bates, 2015) This is a corollary of the previous observation that full equality of opportunity is realised in a Durkheimian spontaneous division of labour in which only natural inequalities are reflected in the inequality of outcomes.

Furthermore, if improved environments lead to a fuller and fuller realisation of genetic potential and thus to higher heritability as less and less of the variation is due to environment, we get a prediction that heritability should vary over time and differ by socioeconomic group. While the twin era studies showed the variation in heritabilities between populations and environments, more recent research has been able to link heritability changes to specific environmental changes, and work is ongoing to better understand the interplay of social inequality and heritability. (Selita & Kovas, 2019)

Interestingly enough, it seems that heritability of education has evolved over time. (Branigan, McCallum, & Freese, 2013) There are suggestions that the introduction of specific education policies may affect heritability (Colodro-Conde, Rijsdijk, Tornero-Gómez, Sánchez-Romera, & Ordoñana, 2015) and some results implying that much of the intergenerational heritability of earnings (Hyytinen, Ilmakunnas, Johansson, & Toivanen, 2019) and education (Engzell & Tropf, 2019) is genetic in a Nordic welfare state context.

Rising of heritability with improved environments also implies the Scarr-Rowe effect, that environmental causes have a stronger effect in lower than in higher socioeconomic groups and heritability is correspondingly higher in higher SES groups. There is some evidence supporting the existence of the Scarr-Rowe Effect. (Fischbein, 1980; Nisbett et al., 2012a) National Collaborative Perinatal Project (NCPP) data has been used to show that in the lowest SES, (shared) environment accounted for almost all of the variation in IQ, while in the highest SES groups almost all the variation was accounted for by genes. The relative weights of genes and environment crossed at middle-class families. (Turkheimer, Haley, Waldron, d'Onofrio, & Gottesman, 2003)

Thus, deficiencies in environment would seem to affect the realisation of genetic potential, with more impoverished environments leading to more limited realisation of the genetic potential. Similar results have been obtained in very recent analyses of Papageorge and Thom (2018) on the interaction of SES and genetics in producing education and labour market outcomes in the United States:

> Our results highlight an important sign change in the interaction between childhood SES and the polygenic score in equations predicting educational attainment. We find that the relationship between the polygenic score and high school completion is weaker among individuals from high-SES backgrounds, while the relationship between the score and college completion is stronger for these individuals. Environments that promote human capital thus appear to be substitutes for genetic endowments in preventing extremely low education levels but may complement these endowments in producing more advanced outcomes.

The model underlying the Scarr-Rowe effect would seem to be supported by a meta-analysis by Tucker-Drob and Bates (2015), which finds it in the US, but fails to find the phenomenon in countries with more developed social policies, namely Western Europe and Australia. This could be due to different social policies in the different countries and could imply that a more generous policy would make the outcomes of the lower socioeconomic groups less a function of the material deprivation of the group and thus lead to a better realisation of biological potential and higher heritability. However, some recent work has failed to replicate these findings (Figlio, Freese, Karbownik, & Roth, 2017), so despite the theoretical and intuitive appeal of the phenomenon, empirical results are not yet conclusive.

Moreover, very recent work in West Africa by Hur and Bates (2019), in a very deprived environment, has produced results indicating similar levels of genetic influences as studies conducted in the affluent West and finding no evidence for the presence of Scarr-Rowe. While the absence of significance of Scarr-Rowe might result from the relatively small effects of background on outcomes, the finding that heritability is on the same level in conditions of severe deprivation does not nicely fit the hypothesis that genetic influences are maximised when environmental conditions improve.

Genetics and environment

Genetics and the environment interact in complex ways. Not only does the effect of genetics on the phenotype always manifest through interaction with the environment, environmental effects are not independent of genetics. Usually, individuals are not assigned to families randomly, so their similarity to their parents does not necessarily arise out of their experiences in their family but may come from their genetic inheritance from their parents. Likewise, the care that they receive from their parents may reflect the genetic dispositions of the parents. Some examples may illustrate the challenges involved.

Generally, the forms of gene-environment correlation (rGE) are divided into three classes – passive, evocative and active – which may all appear at once. The passive form of the correlation occurs when the genetic makeup of parents affects the rearing environment of the children. This nature of nurture has been elegantly illustrated by results showing that non-transmitted alleles are associated with offspring outcomes. (Kong et al., 2018) Thus, not only are the outcomes of an individual affected by her genes, but also genes that they do not possess, but that their parents do. Thus, the affects that those genes have must take place through nurture, taken here to refer to all associated

environmental pathways. Evocative or reactive gene-environment interaction occurs when the environment of the individual responds to heritable behaviours or characteristics of the individual. Examples of this may be found from the heritability of divorce or of the risk of falling victim to bullying. (Jerskey et al., 2010; Veldkamp et al., 2019)

Examples of active gene-environment correlation may be taken directly from intelligence research, which has observed the potency of environmental effects in the Flynn effect, but also made the possibly related, and initially paradoxical, finding that heritability of intelligence rises with age, while the contribution of shared environment correspondingly diminishes. After its initial discovery by Wilson (1978, 1983), the Wilson effect has been rigorously confirmed and is currently generally accepted. The finding was initially paradoxical, as the effect of environment could be expected to increase as exposure to it is extended. It is quite generally accepted to arise because individuals self-select into environments that match their genetic predispositions. This is referred to as the individual multiplier in the Dickinson-Flynn model, and refers to how individuals seek environments that match their genetic dispositions, which augments the variation associated with genetic differences. (Haworth et al., 2009) The interaction of the individual with the environment produces feedback loops by which genetic dispositions and environmental influences form a mutually reinforcing cycle.

Flynn and others have hypothesised that, over the generations, a similar phenomenon produces a social multiplier, where the changing environment may lead to very significant cognitive gains without genetic change. (Dickens & Flynn, 2001; Flynn, 2007) The Dickens-Flynn model remains mostly untested but is consistent with a wide variety of facts about intelligence (Nisbett et al. 2012a). What is clear is that the realisation of genetic dispositions is not environment-invariant and that genetics also affects the environment that individuals end up in.

The hereditarian and environmentalist hypotheses differ in their assumptions as to how significant changes in environment can be, when we measure the significance of the change by its consequences. Indeed, many defences of the hereditarian hypothesis explicitly or implicitly take the existing variation to cover currently available environments and conclude high heritability to entail that the distribution is largely locally non-malleable. (Sesardic, 2005; Herrnstein & Murray, 1994; Jensen, 1969) This makes hopes of environmental cures to e.g. social differences out to be very hopeful indeed. Such analysis is challenged by the very large environmentally driven changes in intelligence illustrated by the Flynn effect. That rapid rise shows that the distribution may change very significantly in a short time, with considerable uncertainty regarding the causally responsible environmental changes. This shows that while environments are not all-powerful, they can change very significantly in causally relevant respects in a very short time. Notably, such changes can occur without us understanding the magnitude of the relevant changes or even reliably identifying the environmental changes responsible for the changes in cognitive achievement. This implies, in turn, that it may be too pessimistic to assume that current distribution reflects biological limits that may be overcome with the introduction of new and yet unknown environments, but that the potential existence of such environments is purely wishful thinking.

7.3 Interpretation of the evidence

The main problem faced by the research is moving from weak genetic explanation to strong explanation. With a growing number of data points, the likelihood of statistically significant false positives

increases, as given significant statistical power, patterns can be found even in randomly generated data. Causality is not a statistical phenomenon, and thus statistical tools are unable to differentiate between causal and non-causal gene-phenotype associations. What weak explanation lacks, and strong explanation has, is a reasonable understanding of the causal biological mechanisms involved in linking base pairs in the DNA to behavioural outcomes. As yet, we are still missing an understanding of the mechanisms. Turkheimer (2012) has argued that behavioural genetics is, with its lack of understanding of the mechanisms, turning into a social science. It is facing the same problem as social science has had: being able to show that certain measured and perhaps plausible causes are associated with specific outcomes, but with no understanding of how that association works.

While Roemer and Trannoy (2016) referred to the problem of lack of genetic information for the estimation of inequality of opportunity, behavioural genetics is experiencing a mirror image of the problem. It can point to very robustly measured associations between genes and outcomes, but it has no answer to the question of whether all the relevant environmental confounders have been identified. We can see what is, but we do not see what needs to be. The difference between the weak and strong biological explanation is thus in very many respects the same as between ill- and well-defined causations. Ill-defined causation needs causation as a fundamental concept because causality is integral to the object of empirical inquiry, but it should not be conflated with well-defined causality. It is true when we accept that interventions change the treated group in all relevant respects. It becomes untrue when we mistake our proxy for a well-defined intervention and forget that we do not have a handle on what all the relevant respects are.

Because of the enormous complexity of reliably identifying strong genetic explanation, popular inferences from genetics have had to rely on assumptions in crucial respects. The hereditarian case for the biological basis of between-group differences has, over the years, quite often been argued on the basis of the heritability estimates, despite the fact that hereditarian authors have also noted that heritability estimates do not describe the malleability of traits. The interpretation has been that, since intelligence is considerably, even 80 %, heritable, the differences between populations are largely genetic and cannot be eliminated by environmental means. The argument has been prominently present in the writings of the most visible proponents of the hereditarian hypothesis (Eysenck, 1973, 1979, 1998; Herrnstein, 1971, 1973; Jensen, 1969, 1998, 2012) – so prominently that it has given rise to accusations that, for example, Hans Eysenck did not understand the concept of heritability. (Colman, 2016)

The reasoning offered in support of the non-malleability of between-group differences in cognitive capacity made an appeal to parsimony, arguing that it is more parsimonious to explain between-group differences using the same mechanisms as within-group ones.

> The default hypothesis states that human individual differences and population differences in heritable behavioral capacities, as products of the evolutionary process in the distant past, are essentially composed of the same stuff, so to speak, controlled by differences in allele frequencies, and that differences in allele frequencies between populations exist for all heritable characteristics, physical or behavioral, in which we find individual differences within populations. (Jensen, 1998)

The point of contention, at least since Lewontin (1970), has been whether or not there are such environmental differences, because treating the group difference as hereditary assumes the environments to be effectively equal. Gottfredson (2013) is one of those who have argued, relatively recently, that when it comes to cognitive development, environments in the 'typical range' are effectively equal.

More recently and generally, it has been argued that genetic causation is easier to observe than environmental causation because one direction of causation is clearly ruled out. '[C]orrelations between a polygenic score and a trait can only be interpreted causally in one direction – from the polygenic score to the trait' and that, thus 'there can be no backward causation from behavior to DNA.' (Allegrini et al., 2018)[21] This has been taken by Plomin (2018) to imply that 'polygenic scores are an exception to the rule that correlations do not imply causation', with the conclusion that

> This correlation means that the inherited DNA differences captured by the polygenic score cause differences between children in their school achievement, in the sense that nothing in our brains, behaviour or environment can change inherited differences in DNA sequence.

This logical inference is, of course, fallacious. In the case of genes and behavioural outcomes, we can be confident that reverse causation is ruled out as a source of spurious correlations. It is less sure that the data generation mechanism does not produce associations that are not causal and can thus be considered spurious.

Even the interplay between environment and genetics can be presented in many different lights. We may emphasise that an environmentally mediated effect on outcomes has very different implications, from both normative and practical perspectives, than a direct genetic effect as the difference

> [It] matters quite a few whits, especially if one should be interested in intervening in the process. Changing behaviour by changing parental attitudes is a decidedly different proposition than tinkering with the ribosomes. (Plomin, DeFries, & Loehlin, 1977)

We may point to the fact that the environmentally mediated effects of genetics – passive effects through parents and active effects through self-selection into environments – emphasise with Plomin (2018) the pervasiveness of genetic influence and that it 'would be a mistake to see gene-environment correlation as inequality, because it is, ultimately, based on genetics.' For our purposes, it is sufficient to note that the difference between these perspectives is mainly one of focus, as one focuses on the possibility of finding well-defined environmental handles to change outcome distributions, while the other focuses on more ultimate causes, and the fact that turning those handles currently depends on genetics.

Here, the crucial issue would seem to be that we are some way away from a strong genetic explanation or from understanding a well-defined causal process that links genetics and phenotypic cognitive outcomes. Thus, the crux of the argument advanced by Plomin (2018), for example, would seem to be the claim that even if the environment has a role and genetic effects may be environmentally mediated, such effects are either uncontrollable or mostly random.

> We can't stop parents from providing correlated nature and nurture to their children unless we adopt children away at birth. We could outlaw selection in schools, but in the classroom it is impossible as well as undesirable for teachers to treat children the same, regardless of their genetic differences. Finally, trying to stop children from actively seeking experiences correlated with their genetic appetites and abilities is futile.

[21] Ironically, even in their study of equality of opportunity, Roemer and Trannoy (2016) recognise reverse causality to be less of a problem, because a child's achievement does not affect parental education.

Currently, there is insufficient evidence as to whether the environmental variables that affect the outcome of the gene-environment interaction are so random as to be beyond policy intervention. Results on the changing heritability over time and the responsiveness of intelligence to environmental interventions would suggest that interventions that can be implemented can have a significant effect. Such results would imply that Plomin may be too gloomy in his perspectives on the tools available to policy.

Causality from observation

Resolving the issue will require significant advances in our understanding of the role of genetics and environment in producing cognitive outcomes. One approach is observational work, trying to solve the challenges of distilling causality from observation through methodological means. Fortunately, it is not entirely hopeless to try to distinguish cases where 'genetic differences between people wind up correlated with phenotypic differences' by chance, from associations that are produced by causal processes arising from the genetic differences in a stronger sense, even if full mechanistic understanding is lacking. Some sources of error are, in the case of inference from GWAS, population stratification, assortative mating and indirect genetic effects. (Selzam et al., 2019; Young, Benonisdottir, Przeworski, & Kong, 2019; Young et al., 2018)

The causal interpretation of any observed association of phenotypic variation with genotypic variation in GWAS is sensitive to hidden population stratification. If, for some reason or another, any given trait is more prevalent in one population or subpopulation than another for entirely environmental reasons, it becomes associated with genetic differences between the two populations. When these two populations are analysed as one, this hidden stratification shows up as an association between genes and traits prevalent in one group and rare in another, even if the two would not be causally associated. (Selzam et al., 2019; Young et al., 2018, 2019) This is a case where an unobserved process in data generation leads to a spurious association. (Pearl, 2009)

GWAS have been shown, with simulated data, to be influenced by the number of principal components that are used to address population stratification. Analyses with real data have also shown that longitude and especially latitude are highly heritable, and this heritability is mostly unaffected even by using a very large number of principal components to correct for population stratification. (Dandine-Roulland et al., 2016) These results seem to be relevant for recent findings of evolutionary selection for height that have found polygenic scores for height increase going from the north through European populations. It has been shown that these results, many of them drawing from the multinational data set of the GIANT consortium, seem to be based on hidden population segregation, as they disappear or become very much attenuated in replication on UK Biobank data. (Berg et al., 2018; Sohail et al., 2019)

Likewise, results from Finland indicate that polygenic scores identify a 'suspiciously large accumulation of geographic differences', which suggests 'bias arising from population structure rather than from a direct genotype-phenotype association.' (Kerminen et al., 2018) The difficulty of distinguishing spurious associations due to population stratification from causal associations that explain differences between populations still makes it difficult to interpret the results of studies that show geographic genetic clustering to be associated with numerous cultural and socio-economic outcomes. (Abdellaoui et al., 2018)

Polygenic scores also perform worse in a replication sample than in the training sample used to generate the PGS. As most scores have been trained on European populations, their performance is worse in non-European, especially in African, populations. (Duncan et al., 2018) This implies that whatever the polygenic score catches, it is at least not the only set of genes that might contribute to a given phenotype. Either the explained phenotypic outcome may arise causally out of different genotypes in different populations, or the polygenic score catches, in the training population, an association that is due to hidden population stratification and thus disappears when the score is used in another population. The robustness of the results may be increased using prediction by a polygenic score in an unrelated population with arguably different unobserved population stratification, as it is plausible that only causally associated alleles contribute to prediction in another population. Such ability to predict in another population increases the reliability of results. (Hill et al., 2018, 2019) However, if the populations are related closely enough, as for example the Scottish and the English might be, then the risk of spurious associations may not be completely removed. On the other side, if there are population-specific variants affecting a specific outcome, such variants will be missed by an analysis that identifies variants predicting an outcome across populations.

Between- and within-family effects

As noted, the challenge is that the main interest of many strands of work relates to between-group differences, making the question of gene-environment correlation impossible to avoid. Recent evidence would suggest that the socioeconomic differences in the frequency of polygenic scores predicting educational attainment are small. Thus, while genetics would predict within-type variation in educational attainment, environmental causes would be much more important than genetics for between-type variation in educational attainment. (Papageorge & Thom, 2018) These results would seem to point in the same direction as earlier results showing more substantial treatment effects with environmental treatments than genetic ones, with a one SD increase in polygenic score giving a 35 % increase in the probability of progressing to next level of education while the same one SD increase in parental education increases progression by 70 %. (Meng-Jung, 2017)

One remedy that has been offered to produce more robust estimates of genetic effects on cognitive outcomes has been to use GWAS to estimate within-family effects because they would be largely free of the confounds due to population stratification or assortative mating. (Young et al., 2018, 2019) Indeed, the recoverability of the anti-Flynn effect in Norway from within-family effects also effectively showed that dysgenic fertility can hardly be to blame for the declining IQ. (Bratsberg & Rogeberg, 2018) Similarly, the case for a causal role of genes associated with educational attainment is strengthened by the finding that siblings with higher polygenic scores for educational attainment tend to be more upwardly mobile than their siblings. (Belsky et al., 2018) This would imply that the scores catch more than the privilege of background.

Recent findings indicate that the ability of polygenic scores to predict individual outcomes is, for cognitive outcomes, significantly better for all variation than for within-family variation. Cognitive outcomes (intelligence, educational achievement) were on average some 60% better predicted between than within families by polygenic score, while polygenic scores predicting non-cognitive traits performed very much the same in both settings. (Selzam et al., 2019) As the difference disappeared when socio-economic status was statistically controlled for, socioeconomic status would seem to have significant predictive power that is not dependent on acting as a proxy for innate ability.

Indeed, recent results using data from the United Kingdom show that, in a multivariable setting, both phenotypic measures of socioeconomic status and polygenic scores for education and intelligence would both seem to significantly predict educational achievement. The purely genetic model explains 18.8 % of the variation of educational achievement, the purely environmental model 30.6 % and the full model including both explains 37 % of the variation, thus performing considerably better than the purely genetic or phenotypic models. Overall, socioeconomic status remains the most powerful predictor in the conditional model. The evidence suggests the presence of environmentally mediated genetic effects and genetically mediated environmental effects. The effects of the genetic model were reduced by 38 % by the introduction of environmental mediators in the model, while genetic mediation explained 18 % of the environmental effects. (Allegrini et al., 2019) These results would imply that both genetics and environments are significant determinants of outcomes and that it would be courageous to claim that socioeconomic status acts as a proxy for genetic dispositions that are causally associated with cognitive outcomes.

The within-family approaches have their practical limitations, however:

> Ideally, GWAS should be performed with parental or sibling genotypes as controls and using models with indirect genetic effects. However, the power of this approach is currently limited because large samples with genotyped siblings and/or parents are uncommon. Furthermore, as only around half of the genetic variation in a population is within family, substantially larger samples of families are required to obtain the same study power as standard GWAS analysis. (Young et al., 2019)

Thus far, the results obtained seem to indicate that the heritability estimates of educational attainment are considerably lower in within-family designs than in between-family methods that catch indirect genetic effects through relatives.

Approaching mechanisms

Other methods that aim to allow for causal interpretation of the role of genetics are related to trying to establish that the genetic variants involved with income, education or intelligence are associated with biological processes plausibly causally associated with the outcome of interest. The plausibility of an inevitable causal role for genetics in educational achievement is more plausible if the genes in question are found to be associated with differences visible in the brain. It would, obviously, be better to understand how the identified genes, SNPs or loci affect the functioning of the brain, but even a robust association is something as it suggests that the identified genetic contribution does not work through affecting skin or hair pigmentation in a system where pigment differences contribute to selection mechanisms in education.

Here, some associations of interest are when the same genes are associated with several behavioural outcomes such as intelligence, education and income, or when genes associated with behavioural outcomes are associated with directly observable biological characteristics. (Donati, Dumontheil, & Meaburn, 2019; Hill et al., 2018, 2019)

Change over time

The fundamental challenge of even very sophisticated work based on observation of the current situation is an inability to extrapolate to what could be. Understandably, another approach to making sense of the possible role of genetics in explaining between-group differences is to observe the evolution of between-group differences over time and the impact of specific interventions. Some of the relevant research on the impact of interventions on intelligence has already been presented above and shall not be reviewed again. The relevant question is related to the transferability of causal evidence to new contexts. Insufficient understanding of how much genotype actually limits cognitive achievement implies that, in many respects, we do not really have good substitutes for trying. (Deaton & Cartwright, 2018)

The discussion on the overall gaps has mostly taken place in the context of racial achievement gaps in intelligence tests and performance in national or international learning outcomes surveys. The discussion has been controversial, in the sense that when evidence has suggested that gaps have decreased, the decrease has been challenged, or if accepted, accepted as taking place in specific subtests and not representing a closing of the gap in *g*. The discussion has also been liberally dosed with accusations of ideological bias and misrepresentation of data. (Colman, 2016; Gottfredson, 2000, 2016; Nisbett et al., 2012a, 2012b) Currently, it seems that strong evidence exists that the racial attainment gap has closed over time, but that progress has not been uniform and the gap has not been fully closed.

For the purposes of the current work, it is sufficient to note only some aspects of the discussion on the evolution of the racial gap in cognitive achievement, which is intimately related to a discussion on the nature of the Flynn effect. At the heart of that discussion is the nature of the Flynn effect in relation to g, the general intelligence factor of intelligence tests. Results indicate that the size of the Flynn effect has been inversely proportional to the g loadings of the different tests. Te Nijenhuis and van der Flier (2013) have interpreted this as a sign that while cognitive performance has clearly changed over time, this change is less significant, as it has happened in areas that have lower g-loadings. Jensen (1998) famously characterised such gains as hollow.

This interpretation has been forcefully rejected by Flynn (2012), who considers that g is effectively a measure of cognitive complexity.[22] He sees that the correlation of the g-factor gives a good case for identifying g with intelligence, but that

> Factor analysis does not capture the dynamic scenario of social priorities altering over time. Thus, g-loadings turn out to be bad guides as to which real-world cognitive skills are merely correlated and which are functionally related. To anticipate, a social change over time such as people putting on scientific spectacles might greatly enhance the ability to classify (similarities) without much affecting everyday vocabulary or fund of general information. Nonetheless all these trends would be of great significance, and to dismiss them as 'hollow' would be a barrier to understanding the cognitive history of our time.

The real point of contention is the relative importance of innate traits in determining cognitive outcomes and their role in group differences. As the Flynn effect has been increasingly recognised – as

[22] We may also note that Gottfredson (2013), one of the most visible proponents of the hereditarian hypothesis, has also, based on Jensen, noted that most *g*-loaded tests are characterised by the complexity of the information processing required.

real and environmental in origin – the discussion has turned to the next ambiguity involved, and it has been correctly noted that any evidence of malleability over time does not amount to evidence of the causes of between-group differences.

> The Flynn effect is predominantly caused by environmental factors, and it is less plausible that these same environmental factors play an important role in explaining group differences in IQ scores. (Nijenhuis & Flier, 2013)

Thus, while the black-white gap seems to have been closing in the United States, it is not clear how long the closing of the gap will continue. Of course, there are still very persistent socioeconomic and cultural differences between racial groups in the United States, so that full convergence would not be expected on the basis of a purely environmental hypothesis. Interestingly enough, recent research on the persistence of racial income disparities also offers suggestive evidence of the possible biological origins of group disparities in cognitive achievement. Results indicate that the persistent income disparities between the white and the black populations in the United States are driven by different gendered patterns of intergenerational mobility in different racial groups. Black Americans have lower levels of upward mobility and higher levels of downward social mobility than whites, but the difference is driven entirely by the male population, as expected values of wages or employment rates between white and black women do not differ once parental income has been controlled for. (Chetty, Hendren, Jones, & Porter, 2018)

The significant gender difference in the black population and the absence of a racial gap for women would seem to strongly favour environmental explanations to the racial gap over genetic ones. After all, black males share much of their genetic stock with black women. As with any empirical research in the social sciences, the results have limitations, and the authors note that the data available does not allow to control for possible racial differences in the selection of women to work. Selection would play a role, even in the presence of racial differences in ability, if high-ability white women and low-ability black women are selected out of work. For now, no credible hypothesis has been presented to support the existence of such differential selection.

Another gap to watch, but one that has been much less present in the debate, is the gap between developed and developing countries. A meta-analysis of 734 studies of Raven's Progressive Matrices test showed convergence over time between developed and developing countries, with the Flynn effect of developing countries outpacing that of the developed ones. (Wongupparaj, Kumari, & Morris, 2015) Full convergence is still a way off, and the future evolution of the gap will be of great interest to any discussion on the malleability of such gaps.

Navigation and uncertainty

Current evidence is insufficient to conclude to what degree phenotypic socioeconomic background is a circumstance and to what degree it entails differences in innate ability. What we should note is the profound analogy that we encounter with the observational study of equality of opportunity. We are engaged in a causal exercise, but identifying well-defined causality is virtually impossible because of the complexity of causal paths, radical multicausality, with myriad interacting causes having small individual effects and thus a pervasive problem of omitted variables.

Unfortunately, the current evidence does not give us much hope for a data source that would allow us to statistically control, even very roughly, for the fact that type may be partially determined by socioeconomic status. Of course, with the current level of understanding of the mechanisms involved, even full genetic data might help us only to a limited degree. 'Just as phenotypic correlations are not self-explanatory, neither are genetic ones.' (Gottfredson, 2011) The question of opportunity is primarily about the possibility of change through environmental manipulation. The pursuit of biological explanations of outcomes by behavioural genetics essentially poses the very same question from a slightly different perspective. It seems unlikely that either field will be ready to provide a well-defined causal account any time soon.

When it comes to the origins – and most of all, malleability – of group differences, the spectrum of positions seems to remain wide. On one side of this discussion we have e.g. Gottfredson and Murray, who see IQ, or at least some IQ differences that are relevant for group differences, as largely genetically determined, while e.g. Nisbett et al. (2012a, 2012b) conclude that 'group differences in IQ are best understood as environmental in origin.'

When assessing the normative implications of the association between genetics and cognitive outcomes, it is crucial to assess the current situation against what could be. If cognitive outcomes were strongly determined by genetics, so that environmental effects could not play a role, or could play only a minimal one, it would be natural to take the distribution of cognitive ability as a fact of nature and thus something to be simply accepted, irrespective of our normative position on whether the genetically determined cognitive outcomes are a just basis for differential welfare or preferably something to be compensated for. (Clark, 2014; Plomin, 2018) In both cases, the normative implications of the situation depend on the ability of individuals or society to influence the distribution of cognitive outcomes. If lower socioeconomic groups or some ethnic groups have lower cognitive outcomes due to strong genetic determination, then resources spent on trying to narrow the cognitive gap between the groups are ultimately futile, but it may yet be a legitimate policy question whether we should redistribute welfare outcomes between individuals to compensate for what is effectively a handicap. But if group differences in cognitive outcomes are not due to strong genetic determination, then the redistribution of opportunity may equalise the between-type differences in cognitive outcomes, which also affects the justification of ex-post redistribution of outcomes.

SECTION IV – EMPIRICAL QUESTIONS AND METHODOLOGY

8 Questions and data

8.1 Questions

The following empirical analyses approach the fundamental questions of education policy from two directions. First, from the perspective of the potential trade-off between equality of opportunity and the level of learning outcomes. Second, from the perspective of the effects of policy on equality of learning outcomes.

As has been emphasised above, we are still far from understanding causally, on the level of well-defined interventions, the intergenerational transmission of social positions and education. Thus, the study of equality of opportunity remains in the unsatisfying situation of studying the effects of complex compound interventions, manifest in social backgrounds, without a specific causal understanding of the relevant aspects of individual circumstances and characteristics that are captured by the indicators of circumstances.

The questions are explored through the keyhole of adult skills surveys, looking for traces of the mechanisms of interest in the distribution of adult skills. The focus of this work is not on what type as a treatment consists of, to better understand the mechanisms of inequality of opportunity. Rather, its focus is on observing the association of observed equality of opportunity with other phenomena of interest, namely the level of learning outcomes and education policy reforms. In brief, the questions pursued are:

1) How are equality of learning opportunity and learning outcomes associated?
2) Are education reforms associated with cohort differences in equality of opportunity?

The first question concerns the predictions of the ability and opportunity hypotheses. The ability hypothesis predicts a negative association between equality of learning opportunity and proficiency, while the opportunity hypothesis predicts the association to be positive. Thus, the question is not so much causal as compositional. It tries to see whether equality of opportunity and excellence are associated as predicted by the ability and opportunity hypotheses. If they are, the interest lies in whether they go together or form a trade-off. This question breaks down into three sub-questions.

1a) Are equality of opportunity and adult skills associated on a systemic level in different parental education types and over different time frames?
1b) What is the dominant direction, if any, of the system-level association of equality of opportunity and proficiency of the high parental education type across cohorts?
1c) Can equality indicators be used to predict proficiency?

The second question approaches causality, by asking whether we can observe an association between education policy reforms and the level of equality of learning opportunity in different cohorts. If education policy has lasting effects, we should be able to link cohort differences in educational experiences to differences in the observed equality of learning opportunity. This question splits into three subquestions:

2a) Can we observe cohort differences in equality of learning opportunity?
2b) When is the relative level of equality of opportunity determined?

2c) Are education reforms associated with observed cohort differences in equality of opportunity?

Any association between cohort differences and education reforms cannot be observed if cohort effects are drowned out by age effects or if cohort differences only stabilise in old age. Thus, only if 2a is answered in the affirmative and 2b indicates an early stabilization of cohort differences, can we possibly link cohort differences observed in adulthood to educational experiences in youth. Even then, the data is insufficient to identify the causal effects of reform but can be suggestive if similar reforms are associated with similar cohort differences.

8.2 Data and measures

Data

The following analyses use data from two large-scale surveys of adult skills – the International Adult Literacy Survey (IALS) and the Programme for the International Assessment of Adult Competencies (PIAAC). IALS measured adult skills in three domains (quantitative literacy, document literacy and prose literacy) and PIAAC in two domains (numeracy and literacy). Quantitative literacy and numeracy are concerned with quantitative reasoning while PIAAC literacy and IALS prose and document literacies are concerned with linguistic reasoning. Due to the nature of the test items in measuring skills in everyday contexts, quantitative reasoning items also require literacy.

In IALS, prose literacy is defined as 'the knowledge and skills needed to understand and use information from texts including editorials, news stories, poems, and fiction'; document literacy as 'the knowledge and skills required to locate and use information contained in various formats, including job applications, payroll forms, transportation schedules, maps, tables, and graphics'; and quantitative literacy as 'the knowledge and skills required to apply arithmetic operations, either alone or sequentially, to numbers embedded in printed materials, such as balancing a checkbook, calculating a tip, completing an order form, or determining the amount of interest on a loan from an advertisement.' (Murray, Kirsch, & Jenkins, 1998)

In PIAAC, literacy is defined as: 'understanding, evaluating, using and engaging with written texts to participate in society, to achieve one's goals, and to develop one's knowledge and potential', and numeracy as ' 'the ability to access, use, interpret and communicate mathematical information and ideas, in order to engage in and manage the mathematical demands of a range of situations in adult life'. PIAAC literacy corresponds to IALS prose and document literacies, while PIAAC numeracy is a broader construct than IALS quantitative literacy, including a broader range of skills, knowledge and situations. (OECD, 2013c)

The three IALS literacies were measured using a pool of 114 items, with each booklet containing about 45 items. (OECD, 2000) The US item pool was 90 items with 64, 66 and 60 items for prose, document and quantitative literacies. (Murray et al., 1998) A two-parametric logistic model from item response theory was used in constructing the scales.

In PIAAC each respondent answered 43 items on three domains: literacy, numeracy, and problem-solving in a technology-rich environment, the last of which has been left out of this work as IALS lacks an even somewhat comparable domain. The computer-based assessment used an item pool of

52 literacy and 52 numeracy items, with 18 literacy and 20 numeracy items used as linking items to the paper-based assessment. The paper-based assessment used an item pool of 24 literacy and 24 numeracy items. PIAAC items were calibrated to link the assessment scale to IALS and ALL (Adult Literacy and Lifeskills) using an item pool of 175 literacy and 74 numeracy items, including IALS and ALL items. For a detailed presentation of definitions and measurement of proficiency, see the technical reports of the surveys. (Kirsch, 2001; Murray et al., 1998; OECD, 2013c)

IALS data was collected in 1994-1998 for 22 jurisdictions. The 1994 data collection included Canada, Germany, Netherlands, Poland, Sweden, Switzerland (German and French) and the United States, all of which completed data collection by February 1995. The 1996 collection included Australia, Belgium (Flanders), Ireland, New Zealand and the United Kingdom. For 1998, data was collected from Chile, the Czech Republic, Denmark, Finland, Hungary, Norway, Portugal, Slovenia and Switzerland (Italian), with Norway and the Czech Republic beginning data collection in November and December of 1997 respectively. (OECD, 2000)

PIAAC data from the first round of the first cycle of PIAAC was collected in 2011-2012 in Australia, Austria, Belgium (Flanders), Canada, Czech Republic, Denmark, Estonia, Finland, France, Germany, Ireland, Italy, Japan, Korea, Netherlands, Norway, Poland, Russian Federation, the Slovak Republic, Spain, Sweden, United Kingdom (England and Northern Ireland) and the United States. The primary data collection window was from the beginning of August 2011 to the end of March 2012. Most jurisdictions began data collection in the autumn of 2011, with only France beginning data collection in 2012. Nine jurisdictions completed their data collection by the end of March 2012, and half of the jurisdictions by mid-April. (Montalvan & Lemay, 2013) The second round of PIAAC data collection took place in 2014-2015, when data was collected in Chile, Greece, Indonesia, Israel, Lithuania, New Zealand, Singapore, Slovenia and Turkey. The third round of data collection took place in 2017 and improved the coverage of the survey mainly in middle-income countries by including Ecuador, Hungary, Kazakhstan, Mexico and Peru.

For PIAAC we have used public use files for all jurisdictions except Australia and Indonesia. Both IALS and PIAAC contain comparable data for 18 jurisdictions, which limits the data to be used for long term analyses. All the jurisdictions in the analyses are presented in ***Table 1.6***.

The object of the analyses is the system-level performance of learning systems vis-à-vis learning outcomes and equality of learning opportunity. The cases of the analysis represent performance in subpopulations defined by parental education, jurisdiction and age group. The jurisdictions used are those used in the IALS and PIAAC data sets. Age group analyses are conducted on several levels of aggregation, ranging from standard 5- and 10-year age bands in IALS and PIAAC to medium- and long-term aggregates, the construction of which is presented below. Use of parental education to define types is presented in the next subsection.

Table 1.6: Jurisdictions and years of data collection

Country	International Adult Literacy Survey (IALS)	Programme for the International Assessment of Adult Competencies (PIAAC)
Austria		2012
Belgium Flanders	1996	2012
Canada English		2012
Canada French		2012
Chile	1998	2015
Cyprus		2015
Czechia	1998	2012
Denmark	1998	2012
Estonia		2012
Ecuador		2017
Finland	1998	2012
France		2012
Germany	1994	2012
Greece		2015
Hungary	1998	2017
Ireland	1996	2012
Israel		2015
Italy	1998	2012
Japan		2012
Kazakhstan		2017
Korea		2012
Lithuania		2015
Mexico		2017
Netherlands	1994	2012
New Zealand	1996	2015
Norway	1998	2012
Peru		2017
Poland	1994	2012
Portugal	1998	
Russia		2012
Singapore		2015
Slovakia		2012
Slovenia	1998	2015
Spain		2012
Sweden	1994	2012
Switzerland French	1994	
Switzerland German	1994	
Switzerland Italian	1998	
Turkey		2015
UK England	1996	2012
UK Northern Ireland	1996	2012
USA	1994	2012

Type as absolute treatment

The following analyses concentrate on the interplay between learning outcomes and equality. The analyses operate on the system level and are concerned with between-type equality of opportunity of achieving learning outcomes. Thus, they focus on the expected outcomes by type, and measure inequality by the difference between best-off and worst-off types. Thus type, defined by background characteristics, is taken as an indicator of individual circumstances or treatment.

Socioeconomic background is measured by parental education, following previous research on equality of opportunity. (Roemer, 1998, Sloat & Willms, 2000) The classification of parental education used is the three-level classification used in both IALS and PIAAC as well as in much other research: 1) neither parent has attained upper secondary education, 2) at least one parent has attained secondary or post-secondary, but not tertiary education and 3) at least one parent has attained tertiary education. These three levels of parental education will be referred to as low, middle and high parental education.

Thus, when measuring equality of learning opportunity, we assume a simple causal model by which the circumstances that are proxied by parental education are environmental and not due to the selection of the parental education groups by innate ability. Thus, the causal model for determination of learning outcomes is supposed to be as presented in ***Diagram 1g*** and the removal of within-type variation of outcomes yields a causal model as presented in ***Diagram 1h.*** The grey line between innate ability and type indicates that, while the effect of innate ability on type has not been eliminated, it is assumed to be effectively immaterial. This assumption is dictated by the insufficiency of the information available on the contribution of genetics to between-type differences in innate ability. Notably lacking is the evidence that would allow us to make country-specific corrections to inequality of opportunity estimates based on country-specific estimates of the level of selection by ability.

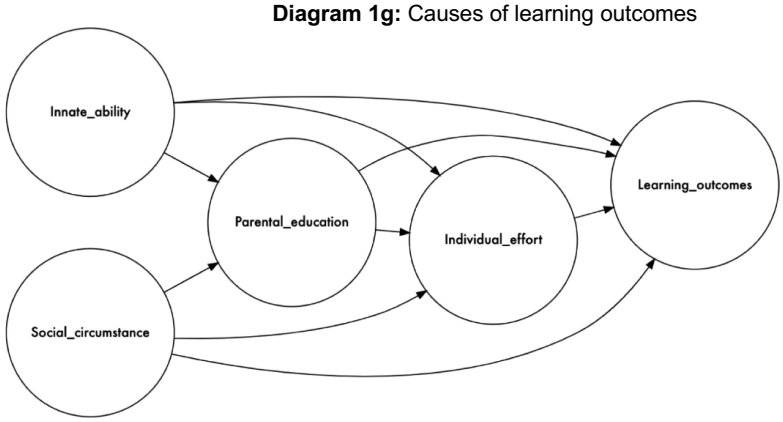
It is also important to note that while this assumption may be false, this is not a shortcoming for the purposes of our current analysis, because the Ability-Opportunity Curve framework does not require the measure of equality of opportunity to be unbiased vis-à-vis innate ability. The crucial part of the assumption that types are not systematically selected by innate ability relates to the implicit assumption that selection by innate ability is uniform in different countries. In the framework of the Ability-Opportunity Curve, the level of between-types selection by innate ability determines the location of the peak, and thus the shape, of the curve. If levels of between-type selection differ markedly in different countries, the countries will not be placed on a unique ability-opportunity curve but on very different ones.

In analytical approaches that have an avowed theoretical preference for treating education as a positional variable, the challenge is estimating reasonable values for, e.g., $10^{th}/90^{th}$ percentiles when the bottom/top categories are relatively large. (Chmielewski, 2019; Reardon, 2011) When the resolution of data around the relevant percentiles is insufficient for direct estimation, the value for the quantile must be extrapolated. This is to be expected, as estimating relative positions becomes difficult if the 7^{th} and 17^{th} percentiles are indistinguishable by direct observation. Reardon and Chmielewski take more than 20 % of the population falling into the top/bottom category to pose a problem for the accuracy of estimation of the 90/10 achievement gap.

Chmielewski's (2019) empirical results also point to the 90/10 gap tending to be underestimated. This is apparent in analysis with parental education and occupation, where relatively large shares of parents fall into the top absolute categories in the recent surveys. She reports that 20% or more of

students fall into the bottom category in 14.8% of cases and into the top category in 38.7% of cases. Unfortunately, she does not report in how many cases either of the conditions is fulfilled. Based on general trends in educational attainment in different countries, it would seem that cases where both the top and the bottom category would surpass 20 %, are quite rare, as the low educational attainment category dwindles before tertiary enrolment becomes massified. (Barro & Lee, 2013) This would imply that there is relatively little overlap in fulfilling the two conditions of having more than 20 % of students in the top/bottom categories. This, in turn, would imply that in some 50 % of cases, either of the conditions is fulfilled. This further implies that a significant portion of the cases analysed suffer from problems in the accuracy of estimating the 90/10 achievement gap and likely underestimate it.

Diagram 1g: Causes of learning outcomes



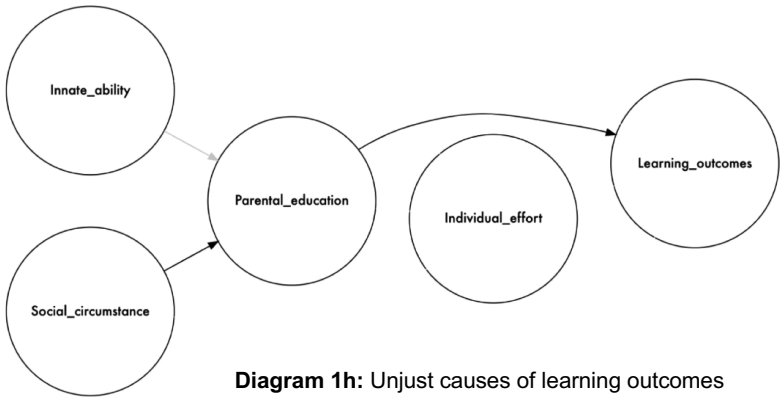


Diagram 1h: Unjust causes of learning outcomes

Centring and standardisation

In the following work, the unit of analysis is a jurisdiction. We are interested in the outcomes of groups rather than individuals. In contrast with earlier research, we centre or standardise equality of learning opportunity and proficiency on the age-group level and construct long-term aggregates from these centred or standardised scores. This procedure is used to take into account the substantial and systematic differences between parental education types and age groups in the levels of proficiency and equality of learning opportunity. ***Table 1.7*** illustrates that there are important differences in achieved outcomes between the parental education types and age groups present in IALS and PIAAC.

The table reports weighted means and standard deviations of mean proficiencies by jurisdiction for all domains and parental education types, thus giving more weight to jurisdiction-cases where a particular parental education type is relatively larger.

Scores are centred or standardised by age group and parental education type in order to make them comparable across data sets, age groups and parental education types. The cases examined are thus effectively divided into separate populations where cases are compared internationally to other cases that share the same age group and parental education type. The average level of comparable cases with the same age group and parental education type in the same data set forms a baseline, relative to which proficiency/equality in an individual jurisdiction is measured.

For some of the analysis, the United States and Chile are outliers for equality of opportunity (USA) or proficiency (Chile), most notably for the high parental education type. Therefore, analyses are conducted both with and without the two countries. The results of the analyses are largely unaffected by their inclusion in or exclusion from the calculation of means and standard deviations for centring and standardisation of scores, so the main analyses reported are performed with Chile and United States included in the centring and standardisation of scores.

The reader may also note that the descriptive statistics of ***Table 1.7*** show that between-jurisdiction variation of achieved proficiency is much larger for the low parental education type than for the middle and especially the high parental education type for all domains and all age bands. This is what we would expect, considering that environmental contributions to outcomes are larger when we are further away from optimal environments. Thus, the general pattern of mean proficiencies by jurisdiction and parental education type is consistent with a model in which we take differences between parental education types to be at least partially environmental in origin.

Table 1.7: Weighted means and standard deviations of jurisdiction proficiencies by domain, parental education, and age band.

	PIAAC Numeracy						PIAAC Literacy					
	Mean			Standard deviation			Mean			Standard deviation		
parental education	low	middle	high	low	middle	high	low	middle	high	low	middle	high
5-year age bands												
16-19	221.9	255.3	276.1	27	19.4	16.8	234.6	264.2	283.1	22.5	15.5	14
20-24	229.1	266.3	286.6	27.2	19.9	17.3	238.6	272.7	291	24.2	17.7	15.9
25-29	233	272.2	291.8	29.2	20.2	19	241.5	277	294.5	25.3	18.1	18
30-34	237.6	273.5	291.3	30.6	18.2	17	243.7	276.6	293.1	27.8	18.2	16.5
35-39	240	273.6	290.9	30.5	19.8	17.4	245.3	276.3	291.9	29.1	18.6	17.3
40-44	239.8	272.8	289.8	30.7	20.1	17.3	245.3	274	289.8	28.1	19.2	16.8
45-49	238.8	270.7	286.9	31.2	18	17.3	244.5	271.9	286.1	27	16.6	16.8
50-54	235.3	267.6	283.1	30.4	18.5	20.3	240.5	267.8	282.5	27	16.9	18
55-59	230.8	262.3	278.2	30.8	17.8	17.4	236.9	263.4	276.5	26.2	16	17.2
60-65	225.8	258.2	268.8	33.2	16.8	19.4	233	259.5	269	27	15.7	17.9
10-year age bands												
16-24	226.2	261.5	281.6	26.9	19.4	16.7	236.9	268.8	287.3	23	16.3	14.7
25-34	235.8	272.9	291.5	29.8	19	17.9	242.8	276.8	293.8	26.5	18	17.1
35-44	240.1	273.2	290.2	30.5	19.8	17.2	245.3	275.2	290.9	28.3	18.7	16.9
45-54	228.7	260.2	273.8	32	17	17.3	242.4	270	284.6	26.7	16.5	16.9
55-65	237.2	269.2	285.2	30.8	17.8	18	234.9	261.4	272.9	26.3	15.5	16.6

Table 1.7 (continued): Weighted means and standard deviations of jurisdiction proficiencies by domain, parental education, and age band.

	IALS Quantitative Literacy						IALS Document Literacy						IALS Prose Literacy					
	Mean			Standard deviation			Mean			Standard deviation			Mean			Standard deviation		
parental education	low	middle	high	low	middle	high	low	middle	high	low	middle	high	low	middle	high	low	middle	high
5-year age bands																		
16-19	258.1	282.4	298.1	24.6	16.2	14.5	261.6	284.8	302.2	24.5	18.9	17.3	259.0	278.1	297.6	20.1	17.4	15.5
20-24	267.4	292.6	309.9	26.5	16.0	11.7	267.3	293.0	313.2	26.9	19.8	13.9	265.6	286.0	306.5	23.7	17.1	14.5
25-29	266.7	298.3	308.8	27.3	17.5	15.3	263.3	296.0	309.0	28.0	20.0	16.4	261.7	288.3	302.5	24.3	18.6	13.3
30-34	265.6	297.2	313.9	28.6	15.7	14.2	261.3	292.5	311.6	29.7	18.9	16.3	259.3	285.0	305.1	26.2	16.7	15.2
35-39	265.3	297.9	307.9	29.1	15.9	18.2	259.5	291.8	303.9	29.6	17.6	19.0	258.0	284.8	299.8	26.6	17.2	18.3
40-44	266.6	292.9	306.9	26.3	17.6	14.7	258.3	285.0	300.9	27.2	20.7	18.5	256.9	279.1	299.5	24.9	21.0	17.7
45-49	264.6	292.0	303.9	24.7	21.1	15.9	255.5	283.7	296.4	25.7	22.2	20.0	254.1	278.6	293.4	23.8	21.6	16.5
50-54	253.8	285.7	296.3	32.5	15.6	16.8	244.8	275.9	290.2	30.5	18.5	20.3	243.8	272.7	288.6	27.0	17.0	17.5
55-59	248.0	279.7	291.5	30.9	16.7	20.8	237.4	268.6	280.9	28.8	21.0	20.0	237.3	264.5	282.4	25.6	21.5	21.4
60-65	237.1	274.8	281.4	31.9	20.6	20.7	224.4	261.4	268.9	32.6	22.7	19.0	224.8	257.5	272.3	29.6	22.3	15.7
10-year age bands																		
16-24	258.5	282.3	298.2	25.2	16.3	14.6	262.1	284.8	302.2	25.2	18.9	17.4	259.4	277.9	297.5	20.8	17.2	15.6
25-34	267.2	293.6	310.1	26.6	15.6	11.0	267.2	294.7	313.5	27.6	19.4	13.2	264.9	286.9	306.7	24.6	16.9	13.7
35-44	268.3	298.3	307.0	27.1	17.8	15.1	264.8	296.6	306.9	28.6	19.8	16.1	262.7	288.4	301.1	24.8	18.1	12.9
45-54	267.7	297.0	314.2	28.8	14.6	14.9	263.3	292.9	311.2	30.0	17.4	15.5	260.6	285.3	305.3	26.4	14.8	14.9
55-65	266.7	298.0	305.7	28.7	15.0	20.9	260.9	292.3	301.3	29.6	16.3	20.3	258.8	285.2	298.2	26.5	15.7	19.7

Proficiency

The first phase of the analysis is deriving average proficiency given survey, domain, parental education type and age group. The unit of analysis is a jurisdiction. To prevent systematic age-group differences in proficiency or equality of opportunity from affecting the results of the analyses, learning outcomes are initially measured on the level of 5- and 10-year age bands. Because the analyses are type-specific, the average proficiencies are derived separately for each parental type.

In the first phase, average learning outcomes are calculated for each case in the analysis. A score is calculated separately for each of the five domains in the two surveys, for all three parental education types and for each age group. Both IALS and PIAAC are surveys using Item Response Theory, so that all surveyed individuals do not answer all items, and thus the proficiency of each individual is estimated rather than measured. A latent regression model is used to produce 10 plausible values, representing achieved proficiency, for each respondent. The practice prefers accurate measurement of proficiency on the level of groups to the accurate measurement of it on the level of the individual. The average outcomes for different groups are calculated in the usual manner from the means of plausible values. It is useful to note, as all the analyses are performed on groups using group means, that group means are thus means of all the plausible values of all individuals in the sample.

With 10-year age bands, this query generates 5 mean proficiencies for the 3 domains and 22 jurisdictions in IALS and 2 domains and 38 jurisdictions in PIAAC for each of the 3 parental education types.[23] Thus the total number of scores generated is 15 per domain for each jurisdiction and a total of 685 scores for each parental education group, for a grand total of 2055 scores in total. Correspondingly, the queries for average proficiencies per 5-year age band produces twice as many cases,

Table 1.8: Structure of the data – cases per parental education type

	IALS quantitative literacy	IALS document literacy	IALS prose literacy	PIAAC literacy	PIAAC numeracy	Total
All						
jurisdictions	22	22	22	38	38	142
5-year age band cases	220	220	220	380	380	1420
10-year age band cases	110	110	110	190	190	710
Comparable data for IALS and PIAAC						
Jurisdictions	18	18	18	18	18	18
5-year age band cases	180	180	180	180	180	900
10-year age band cases	90	90	90	90	90	450

[23] Our regards go to Antero Malin, who conducted the queries for IALS data.

30 per jurisdiction on each domain, for a total of 4110 cases. The structure of the proficiency data is presented in ***Table 1.8***.

Equations

Equation 1: Centring of proficiency scores

$$P_{cjap} = P_{jap} - P\mu_{ap} \quad (1)$$

Equation 2: Standardisation of proficiency

$$P_{zjap} = \frac{P_{jap} - P\mu_{ap}}{P\sigma_{ap}} \quad (2)$$

P_{jap} = raw proficiency score in jurisdiction j in age group a, parental education type p,

$P\mu_{ap}$ = weighted average proficiency of cases in age group a, parental education type p,

$P\sigma_{ap}$ = weighted standard deviation of proficiency in age group a, parental education type p,

P_{cjap} = centred equality proficiency in jurisdiction j in age group a , parental education type p.

P_{zjap} = standard equality proficiency in jurisdiction j in age group a , parental education type p.

Equation 3: Achievement gap as unstandardised inequality of learning opportunity

$$E_{ja} = P_{hja} - P_{lja} \quad (3)$$

Equation 4: Centring of equality of learning opportunity

$$E_{cja} = -E_{ja} - E\mu_{a} \quad (4)$$

Equation 5: Standardisation of equality of learning opportunity

$$E_{zja} = -\frac{E_{ja} - E\mu_{a}}{E\sigma_{a}} \quad (5)$$

P_{hja} = raw proficiency score of a high parental education type in jurisdiction j, age group a,

P_{lja} = raw proficiency score of a low parental group jurisdiction j, age group a,

E_{ja} = achievement gap in jurisdiction j, age group a.

$E\mu_{a}$ = average equality of cases in age group a,

E_{cja} = centered equality score in jurisdiction j in age group a

$E\sigma_{a}$ = standard deviation of equality in age group a,

E_{zja} = standard equality score in jurisdiction j in age group a

Centring and standardisation of scores have been performed in several different ways. The challenge for standardisation of scores is that as neither the averages nor the standard deviations correspond to the general PIAAC or IALS averages or standard deviations calculated for individuals, it is difficult for the reader to interpret the results. Moreover, as there are considerable differences between age groups in the variation of scores by country, the same standard score (Z-score) may correspond to a very different points advantage over the international average in different countries.

The centring of scores preserves the initial scale of IALS and PIAAC, as the proficiency scores are set to achieve 100 points standard deviation in both. Thus, one point is equivalent to $1/100^{th}$ of a standard deviation of IALS or PIAAC, and a centred score of 20 or -20 means that the mean proficiency in a jurisdiction is 0.2 standard deviations above or below the mean of comparable cases that share the same age groups, parental education type and domain.

Many of the analyses have been conducted with both centred and standardised scores. Likewise, analyses have been conducted using scores centred or standardised using the whole sample, as well as standardising so that United States and Chile, which are outliers in some of the analyses, do not affect the mean and/or the standard deviation used in the centring or standardisation. In all cases, the differences in the results are minor. Thus, we prefer centred scores, which are easier to interpret and communicate, as they preserve the original points scale of the two surveys used in the analyses.

When centring and/or standardizing scores, the choice has been made to use weights in calculating the international averages and standard deviations. The cases are weighted by the relative size of the parental education type. Thus, a case where the parental education type A is 60 % of the age group has 3 times the weight of a case where it is 20 % of the population.

The proficiency scores are centred, as presented in ***Equation 1***, and standardised as presented in ***Equation 2***.

Equality of opportunity

The second operation is calculating raw equality of opportunity scores for each case defined by domain, jurisdiction and age group. This means deriving a total of 660 equality of opportunity scores for the 10-year age bands and 1320 scores for the 5-year age bands. Each score is calculated simply as the range of outcomes between the best- and worst-off types, represented by the high and low parental education groups. (***Equation 3***).

The methodology used produces domain-specific measures of equality of learning opportunity. Before analysis, the measure is centred and standardised just as the proficiency scores were. The only difference in centring equality of opportunity and proficiency is that for equality of learning opportunity, the achievement gap is turned into a measure of equality of learning opportunity by using its additive inverse. As the achievement gap is, in essence, a measure of the inequality of learning opportunity, it is turned into a measure of equality of learning opportunity. Thus, a positive/negative equality score denotes a smaller/larger gap than in other jurisdictions on average, and a decrease/increase in the gap between high and low parental education types is referred to as improvement/decrease of equality of learning opportunity.

Unlike proficiency scores, equality of opportunity scores are not parental education type-specific. This means that calculating international averages from national gaps is impossible using weights, and calculation must use unweighted averages and standard deviations. Centring and standardisation of equalities of learning opportunity are presented in ***Equations 4*** and ***5***.

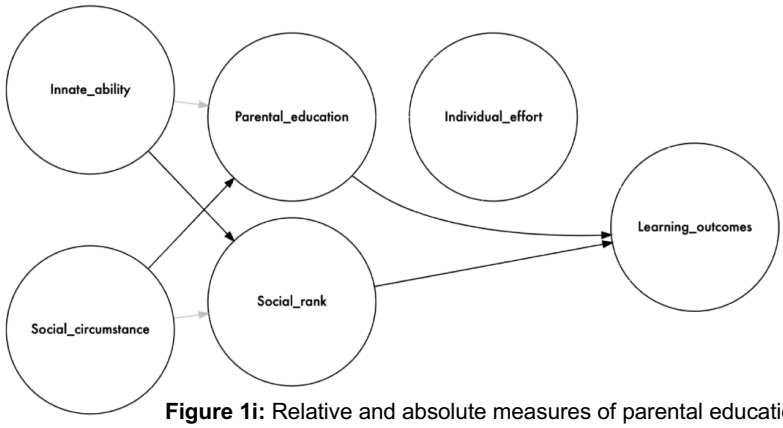


Figure 1i: Relative and absolute measures of parental education

Social rank score - Control indicator for selection and positionality

The choice to measure parental education on an absolute scale implies that parental education is taken to be an absolute rather than relative treatment. Thus, the causally significant aspect of parental education is taken to be the absolute difference between different levels of educational attainment, rather than the relative social position attached to each level of education.

This choice is made for two separate sets of reasons. First, as clarified above, at least some of the effects of education are absolute rather than purely positional. Second, the social rank score is related to the ability hypothesis. If type as a treatment is selection by innate ability, as postulated by the ability hypothesis, then the changing difference in outcomes between different types is explained by changing levels of selection into each type.

As repeatedly emphasised, the methodology above measures equality of opportunity by between-type methodology, assuming that outcome differences associated with circumstances, more specifically socioeconomic background measured by parental education in our case, are due to those circumstances and violate the principle of equality of opportunity. As also noted, this approach assumes that circumstances are independent of effort, which becomes a problematic assumption when innate ability is considered a form of effort and if types have been selected for or are otherwise stratified by innate ability. To allow for the possibility that the treatment effects of parental education are indeed selection effects by innate ability, we thus need a relative measure of the position of the parental education types that would be sensitive to changes in the selection into types.

A proxy measure for selection is necessary to take account of the fact that the composition of the population by parental education has evolved significantly from the oldest cohorts of the IALS, born in the 1930s, to the youngest cohorts of PIAAC, born in late 1990s. Thus, the cohorts of parents range from those born in the early 1900s to those born in the 1970s. High/low parental education represents a very different level of selectivity in the oldest and the youngest cohorts. If the parental education types differ by innate ability due to selection, those differences should be on different levels in different cohorts.

We have explored several means of accounting for the differential level of selectivity by parental education. They share the intuition of the pure ability hypothesis that takes high/low parental education types as the high/low innate ability groups. The smaller the high parental education group, the higher innate ability should be; the smaller the low parental education group, the more select a sample it would be from the bottom of the ability distribution. The supposed key to the level of selectivity is the size of the group. After all, the ability hypothesis supposes that selection into the groups has been unbiased. Even if group size is used as the basis for measuring the level of selection, there are several possible means by which to take the relative size of the type into account.

Instead of directly using the parental group size as a control variable, expecting it to be positively/negatively associated with proficiency in high/low parental education type, we construct a social rank score for each type. This score measures the relative rank of the mean individual of the type. It may be understood as a hypothetical *social rank score,* as the measure captures the idea that relative rank by parental education should be a measure of relative rank by innate ability. If status, as measured by parental education, is a direct manifestation of differences in innate ability, and assuming that innate ability is normally distributed, the high/low parental education types of x/y % represent the highest/lowest x/y % of the innate ability distribution. This allows for a calculation of the mean percentile of a type, which is interpreted to represent the mean ability percentile of the type. For example, if 30 % of a population have low parental education, the mean innate ability of the low parental education type is in the 15th percentile. Correspondingly, if high parental education type represents 30 % of the population, their median innate ability is in the 85th percentile.

Assuming that innate ability is normally distributed, the median percentiles of the high/low parental education types can be expressed in standardised scores. For example, in a normal distribution, the 85th percentile is 1.036 standard deviations above the median. This allows us to calculate a standardised innate ability z-score for each case, knowing only the prevalence of the parental education type in the population.

The distance between the mean innate abilities of the high/low parental education types can be calculated by merely taking the distance between the standard innate ability scores of the types. For example, when both the high and low innate ability groups represent 30 % of the population, the innate ability distance between them is 2.073 standard deviations, as the medians of the high/low types are 1.036 SD above/below the average. The distance between the social rank scores of the best- and worst-off groups is referred to as *social rank distance.*

The relative social rank score may also be understood outside the ability context, in positional terms, as a measure of the positional value of parental education. Thus, when learning outcomes are explained by differences in the absolute treatment of parental education and the relative treatment of social rank (***Figure 1i***), the model contains both absolute and relative measures of parental education.

This modelling choice differs from the approach in some of the previous literature to consider parental education primarily as a positional indicator. (Chmielewski, 2019; Reardon, 2011) The justification for this choice is the fact that because, unlike previous research, the treatment tries to differentiate explicitly between the effects of selection by innate ability and the other aspects of social position as a treatment, it was therefore necessary to construct a measure of parental education that takes into account the changing levels of selection. If parental education had, in addition, been taken to be a primarily positional treatment, we would have needed two differing rationales for measuring positional parental education. Rather than go the way of two positional measures, we have thus used one absolute and one positional measure. The positional measure of social rank then acts as a proxy

for ability selection by parental education and any purely positional benefits with which parental education may be associated.

Social rank scores are calculated for each case, based on the relative size of the parental education type in the data and thus on the relative selectivity of the different types. The social rank score calculation takes relative positions to follow a normal distribution and expresses selectivity by the Z-score calculated from the mean quantile of the parental type.

Income inequality

In the models exploring the association of equality of learning opportunity and the level of learning outcomes, income inequality is also used as a control variable. Income inequality is measured using the Gini coefficients of disposable incomes.

There is a shortage of comparable data, because the Luxemburg Income Study, OECD and the World Bank all lack Gini data for some of the countries in the sample for the years of IALS and PIAAC data collections or even near those years. To avoid comparability issues, even the All the Ginis database of Branco Milanovic has removed the aggregate Gini from the latest updates of the database. (Milanovic, 2019) Thus, comparability issues are not completely avoidable while preserving the full sample of jurisdictions. To use the full sample, Ginis from several different sources have been used in the analyses, trying to limit the number of sources to a minimum. All the sources of Gini are from the All the Ginis database.

The primary source of Ginis is the Luxembourg Income Study for the year of data collection. When LIS data on Gini is not available for the year of IALS or PIAAC data collection, data for the nearest year is used. When the two nearest years are available, each as far from the data collection year as the other, an average of the two is used. Such LIS data is not available for New Zealand (matching neither IALS or PIAAC), Portugal (matching IALS) and several other countries (matching PIAAC). For the IALS jurisdictions and for Belgium, Singapore and Turkey in PIAAC, we use microdata-based World Income Distribution (WYD) database Ginis from the All the Ginis database. For Ecuador, Kazakhstan, Mexico and Peru, we use POVCAL Ginis from All the Ginis database.

For subnational jurisdictions, we use the figures for the national level [England, Northern Ireland, Flanders, Canada (French), Canada (English), Switzerland (German), Switzerland (French), Switzerland (Italian)].

GDP per capita

Following much previous research, such as Natkhov and Kozina (2012), GDP per capita is used as a control variable to account for differences in the level of economic development. We use per capita GDP in purchasing power parity United States 2010 dollars for each country for the year of the IALS and PIAAC data collections, all from World Bank data. For subnational jurisdictions, we use the figures for the national level [England, Northern Ireland, Flanders, Canada (French), Canada (English), Switzerland (German), Switzerland (French), Switzerland (Italian)].

Aggregation

In contrast to earlier research, the analyses on the association of equality of learning opportunity with proficiency are carried out on four different levels of aggregation, taken to represent systemic performance over different time frames. The analyses take account of age and cohort effects in levels of proficiency and equality by centring or standardising both proficiency and learning opportunity scores relative to international averages in the same domain and age group and then constructing longer-term measures from the centred and standardised scores. As the aggregates are constructed from centred scores, the effects of systematic age group differences in e.g. literacy are eliminated from the long-term aggregates.

Table 1.9: Cases by the level of aggregation

	very short-term	short-term	medium-term	long-term
Birth cohorts	5	10	50	63-70
All cases				
cases per parental education type (n)	1420	710	142	36
total cases (n)	4260	2130	426	108
Jurisdictions in both IALS and PIAAC				
cases per parental education type (n)	800	400	80	36
total cases (n)	2400	1200	240	108

Four different time frames are used in the analyses, ranging from the short to the long term. The time frames correspond to four different levels of aggregates, ranging from 5 to 70 birth cohorts. The very-short-term and short-term analyses use IALS and PIAAC data for 5-year and 10-year age bands. The medium-term analyses use the 50 birth-cohort aggregates containing all IALS or PIAAC birth cohorts, respectively. For the long-term analyses, aggregates of IALS and PIAAC data are used, amounting to a total of 63-70 birth cohorts. The number of cohorts used in the long-term analyses differs slightly by jurisdiction, depending on the year of IALS and PIAAC data collections (IALS data was collected in 1994, 1996 or 1998, PIAAC data in 2011-2012, 2014-2015 or 2017). The proficiency and equality of learning opportunity scores of the medium and long-term cases are aggregated from the standardised scores of the short-term cases.

Within each time frame, the analyses are performed on two levels of aggregation across domains. In the primary analyses, each IALS or PIAAC domain is treated as a separate case. In the aggregate analyses, the different domains are combined into an aggregate score representing proficiency in both quantitative and linguistic reasoning.

Aggregation across time frames

For medium-term and long-term analyses, the scores are aggregated from the standardised scores of the short-term cases. Short-term cases (10-year age band) were chosen over the very-short-term (5-year age band) cases for the greater measurement precision in the short term due to higher sample sizes, which provide for a more accurate estimation of proficiency and the achievement gap on the age group level. Aggregation is performed from centred and standardised scores, which avoids the

systematic differences between age groups from driving the results when several cohorts are analysed together. Thus, the scores represent a level of proficiency/equality of learning opportunity against a baseline of the average level of proficiency/equality of learning opportunity of all cases with the same age group and parental education type. This choice should eliminate much of the jurisdiction-invariant age effects, as well as any jurisdiction-invariant cohort effects.

This choice contrasts with earlier research where the whole adult population of a jurisdiction has often been treated as one population despite significant differences in the level and distribution of proficiency between age groups. As we have seen in the section on previous research, this choice has also produced results where some jurisdictions – such as Finland – present low levels of equality of opportunity and high proficiency levels of proficiency in the adult population, but closer inspection shows that within the jurisdiction the age groups with higher levels of equality of opportunity have higher proficiency and vice versa.

The construction of the aggregates across age groups is presented in ***Table 1.10***, which also includes the weights used in aggregation. In the aggregation of equality of learning opportunity scores, each age group is given equal weight. In the aggregation of proficiency scores, which are specific to each parental education type, age groups are weighted by the relative share, measured in percentages, of each parental education type in each age group.

For long term analysis, IALS and PIAAC cases are aggregated, with IALS and PIAAC given equal weight. In this aggregation, IALS prose and document literacies are paired with PIAAC literacy, giving each half the weight of PIAAC literacy. As IALS and PIAAC partially cover the same age groups, their data has an overlap of 31-35 birth cohorts, depending on the year of implementation of IALS. To account for this, cohorts that are represented in both IALS and PIAAC are given half the weight relative to the age groups covered only by IALS or PIAAC. As the method of aggregation narrows standard deviations, the scores are re-standardised in all phases of the aggregation to make

Table 1.10: Cases on four different terms and aggregation of medium and long-term cases

Case	Cohorts	Aggregation method	Original data	Equality weight	Proficiency weight
Very short term	5	-	5-year age band average	-	-
Short term	10	-	10-year age band average	-	-
Medium-term	50	weighted average	10-year age band average	Unweighted (1/5 per age band)	relative size (%) of parental education type
Long term	63/65/67 (IALS 1994-1998)	weighted average	10-year age band average	1/7 per age band.	relative size (%) of parental education type
				Cohorts with overlapping coverage with 1/14 weight (IALS age groups 16-24, 25-34 and 35-44 and PIAAC age groups 35-44, 45-54 and 55-65.)	Cohorts with overlapping coverage with ½ × % weight. (IALS age groups 16-24, 25-34 and 35-44 and PIAAC age groups 35-44, 45-54 and 55-65.)

one point correspond to one standard deviation (SD) in the score distribution of the given level of aggregation.

9 Analyses

The analysis of the association of equality of learning opportunity and proficiency proceeds in three phases.

1) In the first phase, the association of equality of learning opportunity and proficiency is first analysed over four different time frames and on two levels of aggregation across domains. Then the direction of the association over time between the two is explored by predicting proficiency/equality of learning opportunity across age groups and data sets.
2) In the second phase, a model is constructed to predict proficiency by equality of learning opportunity and income inequality and further developed to predict both over time.
3) In the third phase, the analysis focuses on the effects of policy on equality of learning opportunity by analysing the association in time of reforms of compulsory education and equality of learning opportunity discontinuities between birth cohorts.

9.1 Equality of learning opportunity and the level of learning outcomes

Association - Equality of learning opportunity and proficiency in the adult population

The association between equality of learning opportunity and the level of learning outcomes of the adult population is explored using linear regression analysis. In contrast to earlier research, the association is analysed over several levels of aggregation and separately for different parental education types. The analyses use standardised proficiency and equality of learning opportunity scores, the construction of which is presented in the section on measures.

In all regression analyses, the regression is weighted by the relative size (%) of the parental education type. Thus, cases in which the parental education type in question forms a relatively larger share of the age group have a greater weight than cases where it is relatively smaller. This effectively means that countries and (usually younger) age groups with higher levels of parental education have greater weight in regressions of high parental education type, while the countries and (usually older) age groups with lower levels of parental education have higher weight in regressions of low parental education type.

In addition to conducting the analysis for all countries, it is conducted for all countries with the exception of USA and Chile, as well as for the set of 18 countries present both in IALS and PIAAC, and for the countries in both, with the USA and Chile excluded.

Analyses have been conducted for both centred and standardised scores. Both centring and standardisation have been performed both for the entire set of jurisdictions, as well as excluding USA and Chile from calculation of the mean and standard deviation used for centring the scores. These different variations in the treatment of the variable have very minor effects on the results and all the results of the analyses are reported only for the analyses using centred scores.

Analyses with centred scores are preferred over analyses with standardised scores for substantive and communicative reasons. While the level of outcomes quite naturally changes by age, standardisation of scores where the standard distribution of raw scores is different in different age groups can effectively remove some of the relevant information on the variation of between-type differences, which may lead to less predictive power. After all, while one raw score in a large-scale survey of adult skills does not measure a uniform improvement in skills over the entire distribution, the variation in the achievement gap size is small enough to lead us to expect that the raw score differences measure something reasonably invariant in different cases.

More importantly, centred scores are more accessible in reporting the results, as the slope of the association between the equality of learning opportunity and the level of outcomes can be expressed in points of the original scale, which is arguably more intuitive for most readers than standardised scores in a setting where the average and the standard deviation are calculated from jurisdiction averages.

Direction - Equality of learning opportunity and proficiency across cohorts and time

The next part of the first phase of the analysis focuses on the direction of the association between equality of learning opportunity and proficiency. The first step of the analysis shows that equality of learning opportunity is more strongly associated with proficiency on more extended time frames, with a more significant number of cohorts in the analysis, especially for the high parental education type. Thus, equality of learning opportunity and proficiency are associated across age groups and cohorts within jurisdictions.

Interaction across age groups opens up two possibilities. First, aggregating a larger number of observations may simply reduce measurement error, in which case the lower levels of association over shorter terms are due to noise in the measurement, and the long-term results with much stronger association give a better estimate of the association. But if the observation arises simply from measurement error, the interaction across age groups should not have a dominant direction over time. The second option – the interaction of equality of learning opportunity and proficiency across age groups and over time has a dominant direction – is more interesting. An asymmetric and directional association may arise due to both being linked to some common causal process, with either equality of learning opportunity or proficiency causally explaining the other or both reacting to a third cause with differing time delay. While the dominant direction of the association would not establish the predicting variable as the cause of the predicted variable, the fact that equality of learning opportunity could be used to predict proficiency over time, or vice versa, would be valuable for indicator purposes. It is important to note that the following analyses do not attempt to establish a general model for predicting proficiency or equality of learning opportunity over time, nor do they suppose any causal association between the two variables. However, a systematic association between the two across time would mean that we can use one as a reliable indication of the future level of the other. This would be a useful tool to assess the direction of development of the system and could give an early indication on the later effects of recently implemented policies. The analyses in the second leg of the first phase establish how the two measures are associated across age groups and time.

The direction of the association between equality of learning opportunity and proficiency is sought by four different analyses in which equality of learning opportunity of a jurisdiction is predicted across age groups or data sets by proficiency and vice versa. The four analyses focus on the

high parental education type, for which the difference in the strength of association between proficiency and equality of learning opportunity differs most between the terms, which implies that the interaction of equality of learning opportunity across age groups is most evident for this group. Low and middle parental education types are reintroduced in the analyses in the second phase when a model is constructed to predict proficiency.

As all four direction analyses are performed in both directions, with both equality of learning opportunity predicting proficiency and vice versa, any mention of the independent and dependent variables is omitted. Some of the analyses predict the level of proficiency using the level of equality of learning opportunity in an older age group or earlier data set – and vice versa. These analyses are referred to as predicting level by level. In other analyses, change in either of the variables is predicted by change in the other. These analyses are referred to as explaining change with change. All the analyses are performed with short term cases, i.e. 10-year age groups. An overview of the analyses is presented in ***Table 1.11.***

The first two analyses focus on the inter-age-group association between equality of learning opportunity and proficiency within data sets. This means that, e.g., proficiency of young IALS age groups in quantitative literacy is predicted by quantitative literacy equality of learning opportunity in older IALS age groups.

In analyses 2A and 2B, level is predicted by level and change by change within the same data set. In analysis 2A, the level in older age groups is used to predict level in a younger high parental education type. This analysis is performed with several time differences between the predicting and predicted age groups, from a difference of 0 years (predicting within the same age group) to a difference of 40 years (where the level in 55-64-year olds is used to predict level in 16-24-year olds).

In analysis 2B, change between older age groups is used to predict change between younger age groups. In this analysis, a difference between two older age groups is used to explain the difference between two younger age groups. For example, the difference between 55-64-year olds and 45-54-year olds (referred to as 10-year change) is used to predict the same difference (/10-year change) between 35-44-year olds and 25-34-year olds. The difference of 20 years in ages of the predicting and predicted age groups in this example is referred to as the 20-year difference. The analyses are performed for 10-, 20- and 30-year changes and for 0-, 10-, 20- and 30-year differences.

Analyses 2C and 2D predict level by level and change by change across data sets. In analysis 2C, IALS level is used to explain PIAAC level on jurisdiction and age-group levels. On jurisdiction level, we use the level in IALS (1994/1996/1998) to explain the level in PIAAC (2012/2015). One age-group level in each IALS age group (e.g. 25-34-year olds) is used to explain level in the same age group (e.g. 25-34-year olds) in PIAAC. This means that the level in one cohort is used to predict the level in a cohort born and observed 14–21 years later than the predicting cohort. Finally, in analysis 2D, we use differences between older IALS and PIAAC age groups to predict differences between younger IALS and PIAAC age groups, e.g. using the difference between the 45-54-olds of IALS and PIAAC to predict change between the 25-34-years olds of IALS and PIAAC (20-year difference).

Table 1.11: Overview of the direction analyses

	Analysis	Variables: Equality	Variables: Proficiency
Prediction over age groups			
2A	old age groups (10/20/30 birth cohorts) predict young age groups (10/20/30 birth cohorts)	Equality of learning opportunity in age group x	Proficiency in age group y
2B	Change (10/20/30/40 years) between two older adjacent age groups explains change (10/20/30 years) between two younger age groups	Equality of learning opportunity difference between age groups x_1 and x_2	Proficiency difference between age groups y_1 and y_2
Prediction over time			
2C	Level in IALS predicts level in PIAAC (10/50 cohorts)	IALS equality of learning opportunity	PIAAC Proficiency
2D	Change between level between older IALS age group and PIAAC age group explains change younger IALS age group and PIAAC age group.	Equality of learning opportunity difference between IALS and PIAAC in age group y.	Proficiency difference between IALS and PIAAC in age group y.

Predicting proficiency with equality of learning opportunity – the model

After establishing that equality of learning opportunity and proficiency are positively associated, and that in this association equality of learning opportunity is a better predictor of future proficiency than vice versa, we then turn to the construction of a model to predict proficiency using equality indicators. The construction of the model proceeds in phases. First, a contemporary model is constructed to explain proficiency using contemporary indicators of equality of learning opportunity and income inequality, next that model is augmented by the inclusion of indicators of preceding levels of equality of learning opportunity and income inequality. At both stages, the robustness of the model is tested with control indicators.

Construction of the model

In all stages, a model is constructed separately for each parental education type to predict both very short (5 birth cohorts) or short (10 birth cohorts) term proficiency.

The construction of the Coincident models, predicting proficiency using indicators of equality measured at approximately the same time as proficiency, starts with the first stage model, which predicts proficiency using solely medium-term equality of learning opportunity. This model is developed into the Coincident baseline model by adding income inequality into the model. The robustness of the model is tested by the inclusion of control variables (GDP per capita, social rank score and age group) into the Coincident control model. Finally, we test how the inclusion of the short and very short-term equality indicators affects the performance of the model.

After establishing the baseline model explaining proficiency using contemporary equality indicators, it is time to augment the model to perform prediction over time. The first stage model predicts short and very short-term proficiency using contemporary medium-term equality of opportunity, and

Table 1.12: Models predicting proficiency with equality indicators

	PIAAC short term equality of opportunity	PIAAC medium-term equality of opportunity	PIAAC Gini	PIAAC GDP/capita (log)	PIAAC social rank score	PIAAC age group	IALS medium-term equality of opportunity	IALS Gini	IALS GDP/capita (log)	IALS social rank score	IALS age group
Coincident											
First-stage		X									
Baseline		X	X								
Controlled		X	X	X	X	X					
Controlled + short term	X	X	X	X	X	X					
Predictive											
First-stage		X									
Baseline		X	X								
Predictive							X				
Predictive baseline							X	X			
Augmented		X	X				X	X			
Controlled		X	X	X	X	X	X	X	X	X	X
Controlled + short term	X	X	X	X	X	X	X	X	X	X	X

the baseline model adds income inequality, but this time predictions are made only for jurisdictions present in both IALS and PIAAC. This establishes the comparison point to which we can compare the performance of any prediction of PIAAC proficiency using IALS indicators or a combination of IALS and PIAAC indicators.

Predictive and predictive baseline models then predict PIAAC proficiency using the same indicators from IALS, with medium-term equality of opportunity for the predictive model and with income inequality added for the predictive baseline model. This augmented model uses baseline indicators from both IALS and PIAAC, and the Controlled model adds control indicators from both surveys.

An overview of the models is shown in ***Table 1.12.***

Independent variables

The core independent variables of the models consist of the previously calculated indicators of equality of learning opportunity, as well as indicators of income inequality. As the previous analyses will have shown that equality of learning opportunity is more strongly associated with proficiency over the longer terms, the models use two measures for equality of learning opportunity: one for the shorter term and one for the longer term. Primarily we predict very short- and short-term proficiencies using longer-term indicators, but for the most elaborate models, we include a measure of very short / short term equality of learning opportunity. The longer-term equality of learning opportunity is measured using the medium-term (50 birth cohorts) equality of learning opportunity. The values of the medium-term equality of learning opportunity come directly from the medium-term analyses and all 10 very short-term cases and all 5 short term cases covered by medium-term case share the same medium-term equality of learning opportunity.

Assessing the results

For all models the relative contribution of the variables used to the model is analysed using dominance analyses, averaging sequential sums of squares over orderings of regressors. (Grömping, 2015) The coefficients of the variables are presented using the largest model size available.

9.2 Policy – Discontinuities of equality of learning opportunity as evidence of reforms

After demonstrating that equality of learning opportunity and proficiency are associated and that equality of learning opportunity is a better predictor of later proficiency than vice versa, the analysis turns to exploring how equality of learning opportunity differences between cohorts are associated with education policy reforms. All the analyses have been performed with standardised and centred scores as well, but as the results are largely unaffected by centring or standardisation, the reporting stays close to the data and report analyses based on raw achievement gaps in different cohorts and their discontinuities between cohorts.

The feasibility of the project of associating cohort differences with changes in education policy requires certain preconditions. In general, to associate cohort differences with education policy reforms in initial education, 1) policy reforms must have an effect on the distribution of skills, 2) these effects must be relatively permanent over time and 3) they must arise early enough over the life course that they may conceivably arise in initial education. In addition, the sample sizes need to be sufficient to allow for reliable estimates of outcomes for pre- and post-reform cohorts. Equality of learning opportunity is, as used here, a straightforward measure and thus has modest sample size needs, but still requires a reliable estimate of achieved proficiency in both low and high parental education types, so the sample needed is non-negligible. This limits the reliability of estimates on the cohort level.

To understand how such preconditions might be fulfilled, it is useful to recognise that these preconditions essentially describe a simple model with two time periods that differ in the specificity of the environments faced by the cohorts. In the first period, which may be called the setting period, the cohorts face different environments, and thus the relative proficiencies of different types may evolve in different ways relative to each other in different cohorts, which shows up as between-cohort

differences of equality of learning opportunity. In the second time period (the preservation period), the cohorts face a non-cohort specific environment, and thus environmental differences between the cohorts do not lead them to evolve along different paths, with the consequence that the patterns established in the first phase are preserved. It should, of course, be noted that the difference between setting and preservation phases is a stylistic exaggeration, intended to highlight the difference in the different patterns of change over the two periods.

The division into two time periods being conceptual, the key question is whether there is a shift from the first phase to the second, and when that shift occurs, if it does. If different cohorts face different environments over their entire life courses, the relative outcomes of parental education types may evolve over time, and the entire adult life might fall into the setting period. In such a case, any cohort differences observed in adulthood would not necessarily be due to or related to the experiences of the relevant cohorts in their youth. But, if the setting phase ends not long after the end of compulsory education, then distributions even at the age of 70 years will bear the marks of early cohort experiences.

We can hypothesise that the transition from the first to the second time period might happen right after or soon after the end of compulsory education, as the phase of initial education differs from the years in the world of work in terms of the cohort specificity of the environment. In compulsory education, individuals face a highly structured environment that changes by year and cohort, as individuals pass from one grade to another and as curricula and other characteristics of the system are reformed. In the world of work, they arguably face a less cohort-specific environment, which would explain why any pattern of different distributions in different cohorts would be preserved over the life course. Even if the levels of skills were to rise or decline significantly, the relative standings of the cohorts would remain unchanged if these trends were not cohort-specific.

Existing research clearly indicates that education reforms may have effects on the affected cohorts that persist over time. (Böckerman, Hämäläinen, & Uusitalo, 2009; Bratti, Checchi, & De Blasio, 2008; Brunello & Checchi, 2007; Checchi & Flabbi, 2007; Checchi, Ichino, & Rustichini, 1999; Pekkarinen, Uusitalo, & Kerr, 2009; Pekkarinen, Uusitalo, & Pekkala Kerr, 2009; Pekkarinen, Uusitalo, & Pekkala, 2006) This implies that the relatively stable second phase exists and begins soon enough after initial education has ended that any signs of the effects that the reform may have had have not been erased or smudged beyond recognition by subsequent events.

As IALS and PIAAC cover the same age groups in part, their data has an overlap of 31-35 birth cohorts depending on the year of implementation of IALS. We can assess these cohorts with overlapping coverage for similarities in the pattern of discontinuities given by IALS and PIAAC domains. This assessment is based on visually observing the pattern of positive and negative discontinuities for the same cohorts, observed 14-21 years apart in IALS and PIAAC. A match in the pattern of discontinuities provided by the two data sources is taken as an indication that much of the differences are due to cohort effects rather than age effects. Note that this does not preclude significant change over time even after initial education, but simply requires that any changes do not significantly change expected values of outcomes of different types and thus the pattern of equality of learning opportunity across cohorts. This may, of course, be partially or entirely explained by path dependency over life course, so that while adult experiences could, in principle, completely change the ranking of outcomes after end of schooling, the effects of that distribution of outcomes affects the next stages of selection into following phases of education, to the labour market, in the social system etc. so that advantages or disadvantages are perpetuated.

The extent of this stability over time in terms of age groups is also essential. We can imagine a learning system in which the first time period determining the levels of equality of opportunity would continue until age 40 or 70 and only remain stable after that point. Again, existing research implies that changes in initial education are observable in the distribution of outcomes in adulthood and, by extension, that there is a stable phase in the evolution of learning outcome distributions of cohorts. To see if it is observable with crude age band data, the preservation of patterns from IALS to PIAAC in the youngest overlapping cohorts, i.e. the youngest age groups that took part in IALS, would lend credence to an interpretation that precondition 3 has been fulfilled, and the between-type distributions of skills do not change significantly over time after the end of compulsory education.

Many of the results above have been achieved with sufficient adult microdata to allow the effects of the reforms to be observed. This entails precise identification of individuals affected or unaffected by the reforms concerned, and sufficient sample size to allow comparison of the affected with the unaffected, while taking into account e.g. any underlying trends by using differences-in-differences or other similar designs. Unfortunately, the adult skills data available here is not of sufficient sample size to produce reliable birth cohort and parental education type-specific proficiency estimates, and thus comparing the effect of a reform at the margin is out of reach. The main challenge for our approach is the fact that for a given parental education group, the sample size per birth cohort is very limited. As the phenomenon and measure of interest, equality of learning opportunity, understood as the distribution of expected values of learning outcomes against the background of parental education types, requires measurement of expected values of outcomes by type, this leaves us with trying to achieve higher sample size and more accurate measurement by aggregating several cohorts. This sacrifices precision of measurement, as we cannot measure cohorts directly before and directly after the reform.

The chosen method thus attempts to identify discontinuities of interest and possibly associated education policy reforms. It cannot establish that certain changes are due to a specific reform, but may allow us to identify reforms, the effects of which should be more carefully explored with more extensive data. This may be the case, especially when similar reforms appear to be associated with similar changes in later outcomes.

This analysis proceeds in two main phases.

1) In the first phase, we assess how consistent is the picture produced by IALS and PIAAC on the levels of achievement gap / equality of opportunity in different cohorts.
2) In the second, we calculate discontinuities between cohorts and compare the timing of those discontinuities to the timing of known education policy reforms in different jurisdictions.

Equality of opportunity in different cohorts

The analysis begins with the previously calculated equality of learning opportunity using 5- and 10-year age bands for the 18 jurisdictions for which we have data from both IALS and PIAAC. The achievement gap is calculated as presented above in Equation 3.

The age group-specific achievement gaps are matched to the birth years of the age groups. As we are dealing with 5- and 10-year age bands, the birth years used represent the median birth year of each age group. For example, for the PIAAC 2012 jurisdictions, the youngest 5-year age band (16-19-year-olds) is assigned the birth year 1994, as individuals born in 1994 were 18 years old in 2012.

In progressing to older age bands, the matching birth year becomes 5 years earlier with each step, so that the 20-24-year-olds of PIAAC 2012 are matched to birth year 1989, 25-29-year olds to 1984 and so forth. For 10-year age bands, the matching is similar, with the 16-24-year-olds of PIAAC 2012 matched to the birth year 1992, when the 20-year olds of 2012 were born. The four older 10-year age bands are correspondingly matched to birth years 1982, 1972, 1962 and 1952. This matching is obviously done separately for the three rounds of IALS (1994, 1996, 1998) and three rounds of PIAAC (2012, 2015, 2017) so that the addressed birth year of 16-19-year-olds of IALS 1994 (1976) is 23 years earlier than that of the 16-19 years old of PIAAC 2017 (1999).

After the age-group-specific attainment gaps have been matched to birth years to combine data from different surveys, all the surveys are plotted in the same figure, showing attainment gap by birth year as measured by the five domains of the two surveys, both of which have attained full coverage of their respective jurisdictions over a series of three rounds. The pattern arising is inspected to assess 1) the consistency of the overall pattern across time; 2) the presence of individual discrepancies where one survey shows markedly different results than another; and 3) the presence of any age effects in the oldest age groups of IALS or the youngest age groups of PIAAC, where age effects are not revealed by a mismatch in the results produced by the two surveys. In this assessment, special attention is paid to any discrepancy in results between IALS and PIAAC for the youngest age groups of IALS, where age effects can be revealed by a mismatch between IALS and PIAAC. This is assessed, because significant mismatches between IALS and PIAAC for the young age groups of IALS would imply strong age effects, and such effects, especially if they went in different directions in different jurisdictions, would hamper the linking of cohort differences in equality of opportunity to the different experiences of the cohorts during their initial education.

Calculating discontinuities

After establishing that IALS and PIAAC provide a largely coherent picture of the evolution of equality of learning opportunity across cohorts, despite the 14-21 year difference in their administration in the 18 jurisdictions involved, we calculate discontinuities between the cohorts. The aim of this phase is to identify shifts in the level of equality of opportunity to compare those shifts to changes in education policy. In this phase the requirements for the accuracy of the point estimates of proficiency, and the achievement gaps calculated from those point estimates, are higher than in the previous phase, as a similar overall trajectory of proficiency across cohorts may be achieved even if the noise from measurement inaccuracies would make the discontinuities between age groups mainly a function of the noise rather than of actual shifts of equality of opportunity.

In the following analyses, discontinuity refers to a difference in measured equality of learning opportunity between two age groups. The discontinuities are calculated by comparing equality of learning opportunity in two age groups, the older of which represents the pre-reform cohort and the younger the post-reform cohort of a potential reform. The equality of learning opportunity difference between the two age groups is the discontinuity potentially associated with any reform that has taken place between the cohorts. The aim of measuring change from pre- to post-reform cohorts has led to our choice of a 10-year age difference between the compared groups. If the timing of a reform would have been perfect for showing up in analysis using predetermined age bands, a 10-year discontinuity would, because compulsory education has been close to ten years long for all jurisdictions in our sample, compare the first cohorts with experience only from post-reform school to the last cohorts

educated in the pre-reform school. Furthermore, it has been observed in previous research that, for example, the effects of comprehensive school reforms were fully observable after a delay of 8 years. (Borman, Hewes, Overman, & Brown, 2003)

This comparison is made on two levels:

1) 10-year discontinuities, where an age group is compared to one 10 years older (e.g. 25-34-year-olds are compared to 35-44-year-olds) and
2) 5-year discontinuities, where an age group is compared to one 10 years older (e.g. 25-29-year olds are compared to 35-39 year-olds).

Both discontinuities are calculated as

$$D_{yo} = E_{jay} - E_{jao} \quad (6)$$

E_{jay} = equality of learning opportunity in the younger age band, E_{jao} = equality of learning opportunity in the older age band, D = discontinuity between the age groups

The 10-year discontinuities prioritise accuracy in the estimation of the achievement gap, as they achieve larger sample size, while the 5-year discontinuities risk noise and reduced accuracy of measurement of the achievement gap to achieve greater precision in timing by reducing the level of aggregation. It is also important to note that, because we are working with 5-year and 10-year age bands, despite the 10 year difference in the average age of the groups, the age difference between the youngest birth cohorts of the older group and the oldest birth cohorts of the younger group is smaller. For 5-year discontinuities, this difference is 6 years (between e.g. 29 and 35-year-olds), and for 10-year discontinuities, there is no gap between the compared cohorts and the smallest average difference of birth years is only one year (e.g. 34 and 35 year-olds). Thus, in the 10-year discontinuities, there is a considerable overlap in the schooling experience of the compared groups. Thus, it is to be expected that the magnitude of the discontinuities is reduced in analyses with 10-year age bands. Additionally, as it obviously takes more time for any changes to affect the average of a 10-year age band than a 5-year one, the 10-year discontinuities would show any changes at a later point in time than the more volatile and quick-reacting 5-year discontinuities.

As our priority is in identifying changes and assessing their timing, we thus prioritise the 5-year discontinuities over the 10-year ones when we match discontinuities between cohorts to changes in education policy. Even with the 5-year age bands, where the overlap of the groups compared is smaller in terms of their experiences of compulsory schooling, the time spent in the education system overlaps substantially, especially if we take into account time spent in upper secondary and tertiary education.

Timing the discontinuity

After calculating discontinuities between cohorts, they are timed as well. The discontinuities, i.e. differences between close cohorts, can be presented in various ways. As the discontinuities are essentially cohort differences, they might be timed by the birth year of the respective cohort. The drawback of this method of presentation is that timing of the affected/unaffected cohorts differs from

timing of the relevant reform. Thus, a discontinuity between the 16-19 and 25-29-year-olds of PIAAC 2015, matched above to birth years 1997 and 1987, would be placed in the late 1980s or very early 1990s. Here it would be important to keep in mind that a discontinuity timed in the early 1990s could well be a consequence of a reform that took place in the early 2000s or even somewhat later.

We have chosen to present the timing of the discontinuities in a way that would directly link the discontinuities as closely as possible to any reforms they might be associated with. As we seek to associate the discontinuities with policy changes in compulsory schooling, and for many of the countries compulsory schooling ended by the age of 16, we time the discontinuities in the year when the youngest birth cohort of the older age group (representing the pre-reform cohorts) used to calculate the discontinuity. The youngest possibly unaffected cohort in our example comparing the 16-19 and 25-29-year-olds of PIAAC 2015, matched above to birth years 1997 and 1987, are the 25-year olds of 2015, born in 1990. Thus, we time the discontinuity between the two age groups in 2005 when the group was very close to finishing compulsory education. Correspondingly, the difference between the 16-24-year olds and the 25-34-year olds of PIAAC 2015 is timed as having occurred in 2005.

Obviously, this method is approximate, as e.g. the length of compulsory education has varied and is different in different countries. The timing method adopted tries to avoid any temptation to fit the timing of discontinuities to reforms by means of jurisdiction-specific timing methods, as well as to be comparable with the timing of discontinuities that may be found using international large scale surveys of youth skills – such as PISA, TIMSS or PIRLS.

It is important to note that because this method of timing by design includes the delay in the full effects of any reform, it produces timings that are very close to and sometimes even precede known education reforms.

Presentation of discontinuities in the figures and in text

The discontinuities are presented visually by jurisdiction. Both IALS and PIAAC discontinuities are included in the same figures to allow for easy comparison of the results from the two data sets. In these figures, the timing of the discontinuity is presented as calculated above, which lines IALS cohorts up with corresponding PIAAC cohorts. The domains used to calculate the discontinuities are not identified in the main figures to avoid unnecessary complexity that might make interpretation of the figures more complicated than necessary.

In the analysis, the timing of the discontinuities is treated as approximate, despite the appeal of referring to exact matches of a timed discontinuity and an implemented reform as evidence of the exactness of the method. Due to the method of calculation, the timing of the discontinuities is approximate, and the description of the results tries to reflect this fact.

Matching discontinuities to reforms

After observing the permanence of the pattern of equality of learning opportunity discontinuities, the observed pattern of discontinuities is compared to the timing of reforms of education policy, mainly of compulsory education (question 7). The reforms are mostly identified using two compendia of education reforms, and their data on the timing of reforms and the affected cohorts. (Fort, 2006; Garrouste, 2010)

In the matching of discontinuities to reforms, the emphasis is on significant education policy reforms affecting the length of compulsory schooling, the age of tracking/streaming etc. Some additional literature is sought when the compendia do not provide sufficient information to find education policy changes whose timing would correspond to the observed discontinuities. In this stage, some 30 education policy reforms are identified across 10 jurisdictions. The analysis does not focus on individual reforms but instead, after first identifying the reforms associated with noticeable discontinuities, on commonalities between the observed reforms across jurisdictions.

SECTION V – RESULTS

10 Equality of learning opportunity and level of outcomes

The first section of the empirical analyses presents the results of the co-occurrence and interplay of equality, especially equality of learning opportunity, and the level of learning outcomes. It does not attempt to establish a causal relationship between the two, but rather to contribute to a better description of their evolution over time and space. As presented above, the coevolution of the two has been a perennial, even fundamental, question not only of education policy and sociological research into education but – in the broader context – of political theory and philosophy. There is hardly anything more fundamental to questions of social structures, hierarchy and power than the question of innate ability and personal privilege as explanations of the lives we lead and the positions we achieve.

The following sections show that equality of learning opportunity and proficiency are positively associated on a jurisdiction level in all parental education types and from very short to long terms. The association is stronger for groups with lower parental education and on higher levels of aggregation across age groups and/or domains, and longer-term equality indicators are able to explain a large share of the variation of learning outcomes, which also allows equality indicators to predict proficiency over time.

10.1 Equality of learning opportunity and proficiency are positively associated

The first results on the association between equality of learning opportunity and the level of learning outcomes are presented in ***Table 2.1a*** and ***Figure 2.1a***. Both the table and the figure collect all three parental education types and the four timeframes together to better illustrate the differences between the types and the timescales. In total, the multiplot collects information on 12 different regressions for a grand total of 6924 data points. The timescale becomes longer when progressing upward from the bottom of the figure and the parental education type becomes higher when progressing from left to right. As the number of cases in each section of the multiplot depends on the level of aggregation, the size of the points varies as well, both to allow observation of the points in the long term figure and to allow at least some observation of the location of the points in the very short and short term analyses with very much higher numbers of points. In each section of the multiplot, equality of opportunity is plotted on the horizontal axis and the level of learning outcomes on the vertical axis.

Equality of learning opportunity and proficiency are, in general, positively associated in all parental groups, but the strength and impact of the association depend on parental education and time frame. For the low parental education type, the association between equality of learning opportunity and proficiency is robust, with a high coefficient of determination, and the impact of the association very high, with a steep slope of association. For the middle parental education type, the impact and strength of the association are weaker, and the difference between time frames in the strength and impact of association is more substantial. The high parental education type shows the lowest levels of impact and strength of the association.

The association is statistically significant for the low and middle parental education types on all time frames. For the high parental education type, the association is statistically significant in the short, medium and long terms, but not in the very short term. On all time frames, the lower the level of parental education, the stronger is the association between equality of learning opportunity and

proficiency. The difference between parental education types is especially marked in the very-short-term and short-term analyses, where variation of equality of learning opportunity explains 35.0 % and 43.7 % of the variation of proficiency for the low parental education type but is statistically insignificant for the very-short-term high parental education type and only explains 0.8 % of the variation in the short-term analyses.

As is to be expected, the lower the parental education type, the steeper the slope of the association. For all types, the slope is steeper the longer the time frame in question. While, for the low parental education type, an achievement gap one point smaller is associated with a 0.88 point higher proficiency in the very short term, in the long term it is associated with 1.82 additional points of proficiency. For the high parental education group, for which the between-country variance of gaps is much smaller, a one-point narrowing of the achievement gap only brings 0.03 points of proficiency in the very short term (which is without statistical significance), but a not-inconsiderable 0.43 additional points in the long term.

In terms of the coefficient of determination, the differences between the time frames are most pronounced for the high parental education type, for which the association of equality of opportunity with proficiency is not even statistically significant in the very short run. For the low parental education type, variation of equality of opportunity explains 35.0 % of the variation of proficiency in the very short term and 59.6 % in the long term; for the high parental education group, the change is from a statistically insignificant R^2 of 0.001 in the very short term to a more robust 23.2 % in the long term.

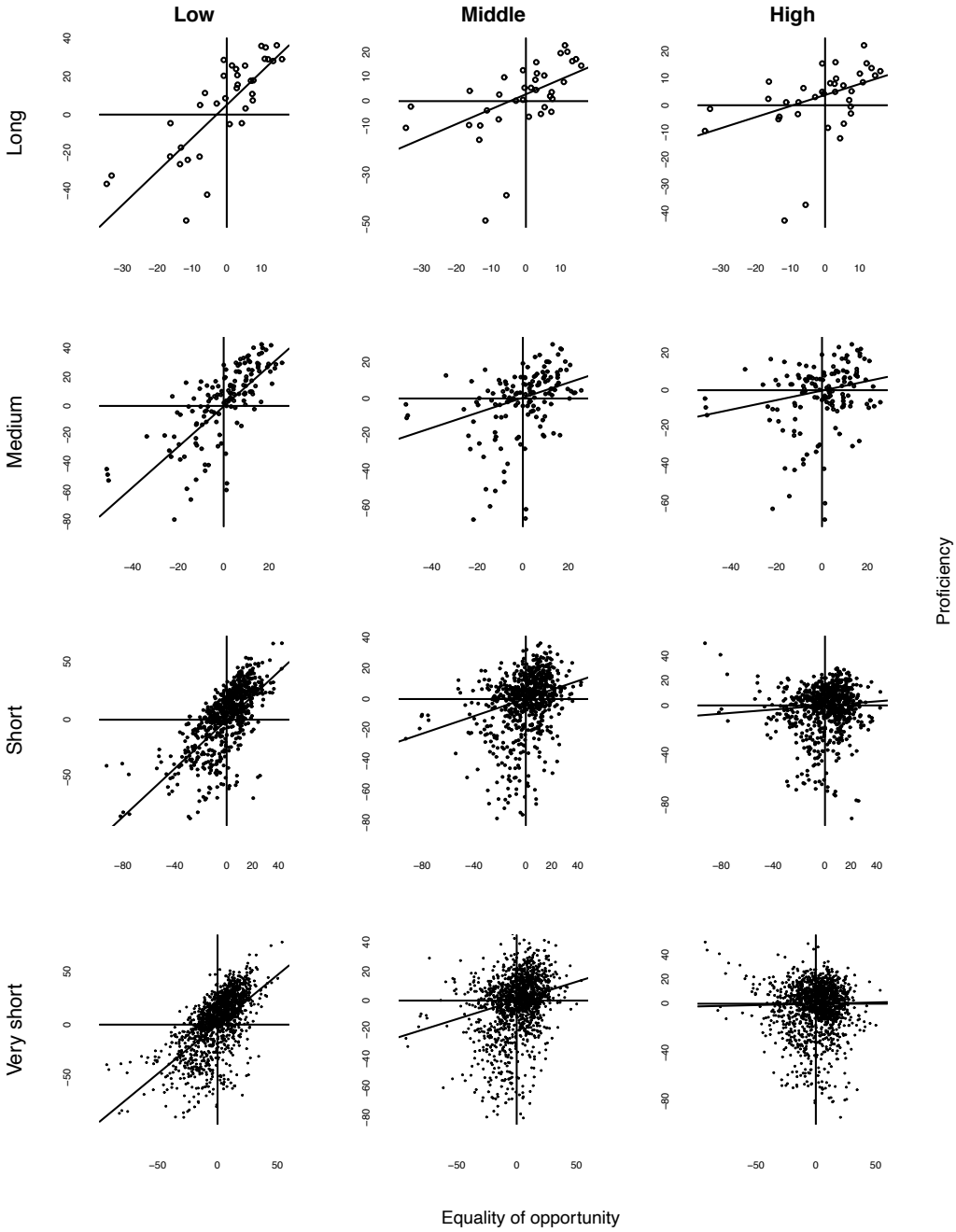
In long-term analysis, the USA and Chile are relative outliers, especially when it comes to the high parental education type. The former has a relatively high proficiency score relative to equality scores below -30, and the latter has relatively low proficiency. In addition, the difference between the very short to medium terms and the long term is affected by the fact that the long-term analysis is based on a subset of 18 jurisdictions that took part in both IALS and PIAAC, as against the 43 jurisdictions that took part in either IALS or PIAAC and were used in the very short to medium-term analysis. Thus, the difference between time frames from very short to medium terms is driven by different levels of aggregation, but from the medium to the long term the difference is also affected by a difference in the set of jurisdictions analysed. This shortcoming can be addressed by separately analysing only the jurisdictions that have taken part in both IALS and PIAAC.

Table 2.1a: Equality of learning opportunity and proficiency are associated (all jurisdictions, centred scores)

parental education	term	n	slope	SE	R2	p
low	**very short**	1420	0.93	0.03	0.350	0.000 ***
	short	710	1.05	0.04	0.437	0.000 ***
	medium	142	1.39	0.12	0.512	0.000 ***
	long	36	1.74	0.25	0.596	0.000 ***
middle	**very short**	1420	0.26	0.03	0.061	0.000 ***
	Short	710	0.28	0.04	0.082	0.000 ***
	medium	142	0.41	0.09	0.133	0.000 ***
	Long	36	0.60	0.16	0.306	0.000 ***
high	**very short**	1420	0.02	0.02	0.001	0.388 -
	Short	710	0.08	0.03	0.008	0.017 *
	medium	142	0.25	0.08	0.064	0.002 **
	Long	36	0.41	0.13	0.232	0.003 **

Note:
* = p < 0.05
** = p < 0.01
*** = p < 0.001

Figure 2.1a: Equality of learning opportunity and proficiency by the level of parental education and length of term (all jurisdictions)



The analysis of countries present in both IALS and PIAAC. ***(Table 2.1b, Figure 2.1b***) is obviously the same as before for the long term, but the overall picture is also very similar to that for the analysis of all jurisdictions. If anything, the association between equality of opportunity and the level of learning outcomes grows stronger.

The most notable difference is for the high parental education type, for which the slope of the association is initially much steeper than reported above (0.07 vs 0.02), and both the slope of the association and the coefficient of determination increase more gradually from the shorter term to the longer term. Unlike in the analysis of all jurisdictions, while equality of opportunity explains a very small share of the variation of proficiency of the high parental education type in the very short run, the association between equality of opportunity and proficiency is statistically significant on all time frames.

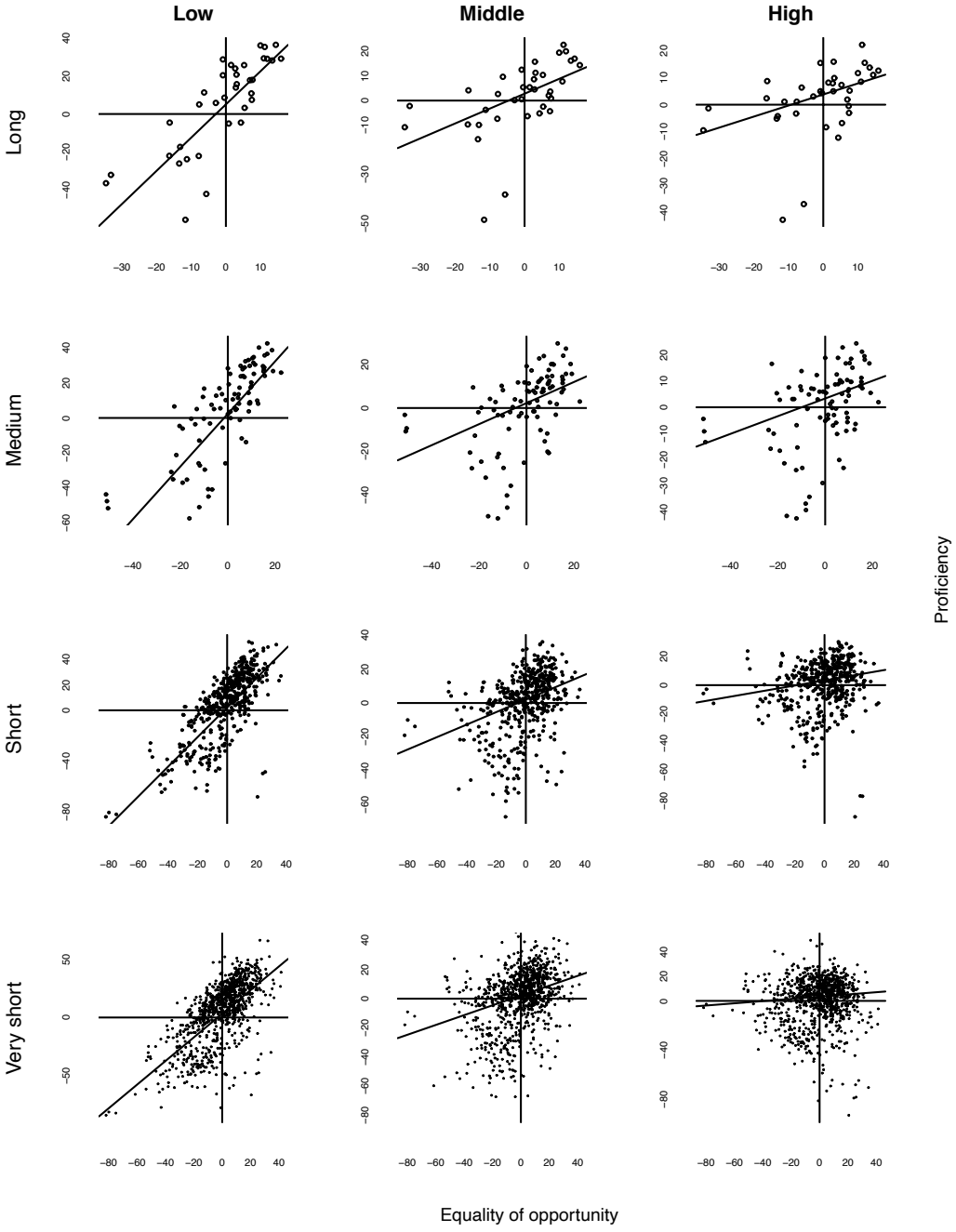
Even in this setting we still need to be concerned about the effect of the two long-term outliers, the USA and Chile. While they are reasonably close to the main group of countries even in the long term with the low parental education type, they are clearly outliers with the high parental education group, especially Chile with its very low proficiency. Note that this status as outliers led to the decision to exclude these two jurisdictions from the calculations involved in centring the scores used for the analyses. The two jurisdictions may also be removed from the analyses to make sure that they do not have an undue influence on the results of the analyses.

Table 2.1b: Equality of learning opportunity and proficiency are associated (jurisdictions in IALS and PIAAC, centred scores)

parental education	term	n	slope	SE	R2	p
low	very short	900	1.01	0.04	0.380	0.000 ***
	short	450	1.17	0.05	0.505	0.000 ***
	medium	90	1.50	0.13	0.620	0.000 ***
	long	36	1.74	0.25	0.596	0.000 ***
middle	very short	900	0.34	0.03	0.117	0.000 ***
	short	450	0.37	0.04	0.157	0.000 ***
	medium	90	0.48	0.10	0.226	0.000 ***
	long	36	0.60	0.16	0.306	0.000 ***
high	very short	900	0.09	0.03	0.011	0.001 **
	short	450	0.18	0.04	0.049	0.000 ***
	medium	90	0.34	0.08	0.170	0.000 ***
	long	36	0.41	0.13	0.232	0.003 **

Note:
* = p < 0.05
** = p < 0.01
*** = p < 0.001

Figure 2.1b: Equality of learning opportunity and proficiency by the level of parental education and length of term *(jurisdictions in IALS and PIAAC)*



Removal of the USA and Chile, both of which are outliers, has a fairly modest effect on the results. (***Table 2.1c, Figure 2.1c***) For the low parental education type, the slopes are somewhat flatter and the coefficients of determination larger than before. For the middle parental education type, the slopes are somewhat steeper and the coefficients of determination somewhat higher than those reported above, with a one-point decrease in the achievement gap being associated with an increase of proficiency by 0.85 points on the long run.

For the high parental education type, the difference between time frames is more pronounced, with no statistically significant association between equality of opportunity and outcomes in the very short or short runs. In the long run, a not inconsiderable 0.37 point increase of proficiency is associated with narrowing the achievement gap by 1 point.

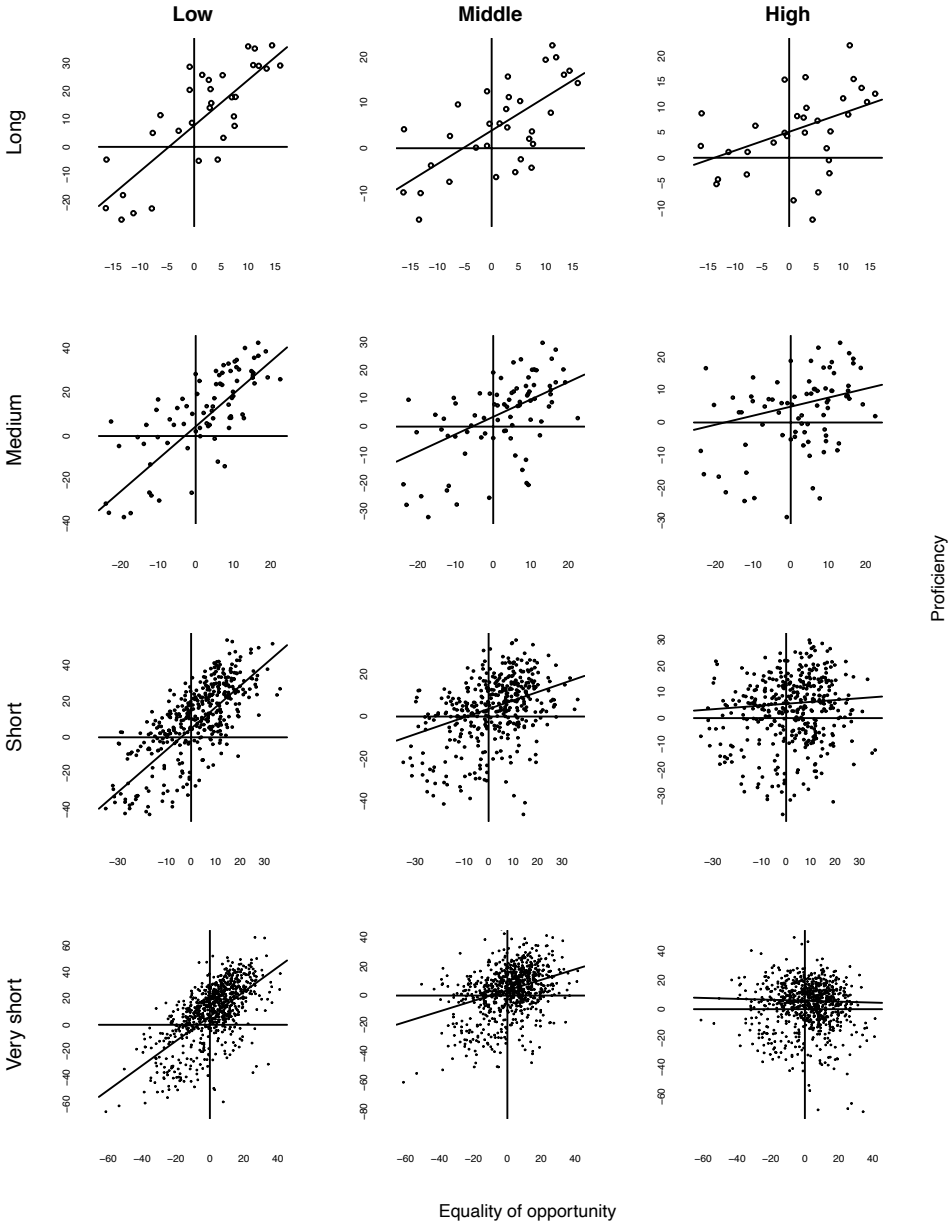
The overall pattern of the results is very much the same when using standardised proficiency and equality of opportunity scores.

Table 2.1c: Equality of learning opportunity and proficiency are associated (jurisdictions in IALS and PIAAC, without Chile and United States, centred scores)

parental education	term	n	slope	SE	R2	p
low	very short	800	0.93	0.04	0.373	0.000 ***
	short	400	1.18	0.05	0.540	0.000 ***
	medium	80	1.49	0.13	0.617	0.000 ***
	long	32	1.64	0.22	0.640	0.000 ***
middle	very short	800	0.36	0.03	0.143	0.000 ***
	short	400	0.40	0.04	0.167	0.000 ***
	medium	80	0.62	0.11	0.299	0.000 ***
	long	32	0.74	0.14	0.495	0.000 ***
high	very short	800	-0.03	0.03	0.001	0.294 -
	short	400	0.07	0.04	0.007	0.103 -
	medium	80	0.28	0.10	0.083	0.010 **
	long	32	0.37	0.13	0.202	0.010 **

Note:
* = p < 0.05
** = p < 0.01
*** = p < 0.001

Figure 2.1c: Equality of learning opportunity and proficiency by the level of parental education and length of term (jurisdictions in IALS and PIAAC, without Chile and United States)



Income inequality and proficiency

Later in the analysis, we construct a model that predicts proficiency using indicators of equality. To understand the role of income inequality in the model, it is useful to briefly see how income inequality is associated with equality of learning opportunity and with proficiency.

The Gini coefficient is negatively and statistically significantly associated with proficiency on all time frames and for all parental education types (***Table 2.2a, Figure 2.2a***), with the strongest association with the low parental education types, and the weakest with the middle parental education type. In the long term, we may observe that there are very few cases where above/below average proficiency is achieved in the long term with very unequal/equal income distribution.

It should also be noted that the Gini coefficient explains the proficiency of the high parental education type in the very short and short terms markedly better than equality of learning opportunity. In addition, the slope of the association between income inequality and proficiency does not change between timeframes as it does for equality of learning opportunity and proficiency, especially for the high parental education type.

Here it is useful to remember that we do not have an age-specific measure of income inequality, but one that applies to the whole adult population. Thus, in our framework, it is a medium-term indicator, in contrast to the more specifically age-group- or cohort-specific indicators.

Table 2.2a: Income inequality and adult proficiency are negatively associated

parental education	term	n	slope	SE	R2	p
low	very short	1420	-2.72	0.08	0.489	0.000 ***
	short	710	-2.80	0.11	0.471	0.000 ***
	medium	142	-2.80	0.23	0.514	0.000 ***
	long	36	-2.74	0.43	0.550	0.000 ***
middle	very short	1420	-1.55	0.07	0.244	0.000 ***
	short	710	-1.64	0.09	0.296	0.000 ***
	medium	142	-1.64	0.20	0.328	0.000 ***
	long	36	-1.57	0.27	0.504	0.000 ***
high	very short	1420	-1.55	0.07	0.274	0.000 ***
	short	710	-1.58	0.09	0.307	0.000 ***
	medium	142	-1.58	0.18	0.365	0.000 ***
	long	36	-1.47	0.20	0.624	0.000 ***

Note:
* = p < 0.05
** = p < 0.01
*** = p < 0.001

Figure 2.2a: Income inequality and proficiency by the level of parental education and length of term

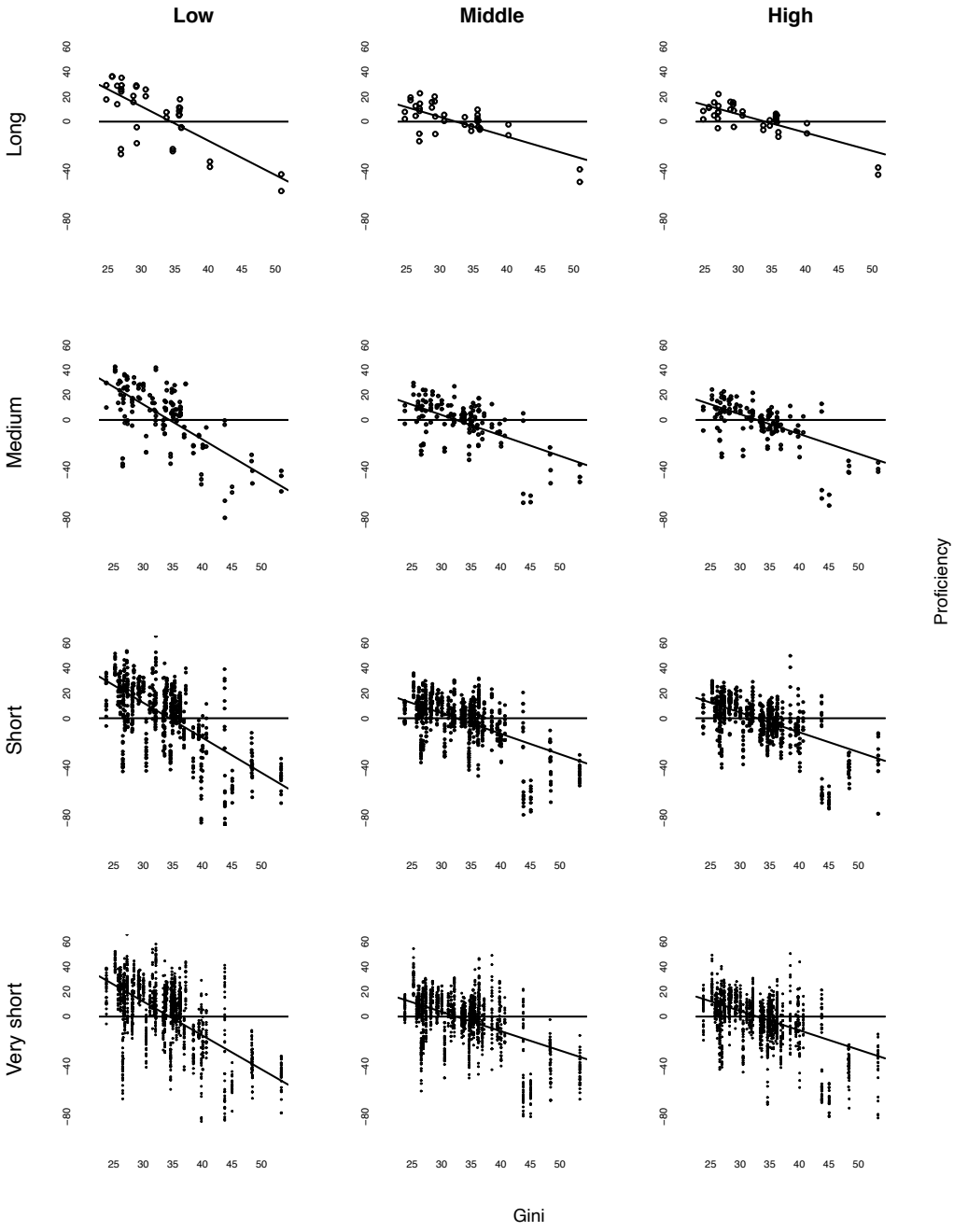


Table 2.2b: Income inequality and equality of learning opportunity are negatively associated, and weighting affects the association

weighting parental education	term	n	slope	SE	R2	p
low	very short	1420	-0.73	0.07	0.077	0.000 ***
	short	710	-0.83	0.09	0.103	0.000 ***
	medium	142	-0.77	0.12	0.148	0.000 ***
	long	36	-0.68	0.26	0.170	0.012 *
middle	very short	1420	-0.61	0.08	0.040	0.000 ***
	short	710	-0.77	0.11	0.064	0.000 ***
	medium	142	-0.81	0.20	0.102	0.000 ***
	long	36	-0.97	0.30	0.230	0.003 **
high	very short	1420	-0.81	0.08	0.070	0.000 ***
	short	710	-0.93	0.11	0.091	0.000 ***
	medium	142	-0.80	0.18	0.127	0.000 ***
	long	36	-1.22	0.31	0.310	0.000 ***
unweighted	very short	1420	-0.70	0.07	0.063	0.000 ***
	short	710	-0.80	0.10	0.083	0.000 ***
	medium	142	-0.80	0.18	0.127	0.000 ***
	long	36	-0.85	0.29	0.207	0.005 **

Note:
* = p < 0.05
** = p < 0.01
*** = p < 0.001

Income inequality is also associated with inequality of learning opportunity. Here, it is useful to note that, as all of our linear regression analyses are weighted to give more weight to cases where the share of e.g. low parental education type is higher, a weighting scheme must also be chosen for the association of inequality of opportunity and income inequality. The importance of the weighting scheme is demonstrated in ***Table 2.2b***, where equality of learning opportunity is predicted by the Gini coefficient. It can be noted that the Gini is more strongly associated with inequality of opportunity when the cases are weighted by the share of the high parental education type than with other types or with unweighted regression.

Thus, our two measures of equality, equality of learning opportunity and income inequality, are closely linked. Consistently with existing literature on, e.g., the Great Gatsby curve, we can observe that equality of learning opportunity is negatively associated with income inequality. But even though the two are associated, the link between them is not very tight. While income inequality does very

well in predicting proficiency on the short and very short terms, it does considerably worse at predicting inequality of opportunity on those timescales. Even in the medium and long terms, income inequality explains a relatively small share of the variation of inequality of opportunity, especially compared to predictions of proficiency.

Interestingly enough, even countries with very unequal income distribution, namely Chile, fall very close to the regression line when proficiency is regressed against income inequality as in ***Figure 2.2a***. This is of interest also because Chile has been an outlier in some previous analyses on the high parental education type by virtue of having very low learning outcomes relative to the level of inequality of learning opportunity.

Main takeaways

There are three main takeaways from the previous analyses.

First, larger inequalities of learning opportunity and of income seem to be very consistently associated with lower proficiency for all parental education groups and on all time scales. For the high parental education group, the positive association is sometimes very weak or non-existent in the very short and short terms but always present in the medium to long term. The impact of the association is non-negligible, with the proficiency of even the best-off type, the high parental education type, improving in the long term by 0.37-0.41 points with every point that the achievement gap is narrowed, depending on the jurisdictions included in the analyses. In the long term, variation of equality of learning opportunity also explains some 0.202-0.232 of the variation of proficiency. The negative association of income inequality with proficiency is considerably stronger, most markedly for the high parental education type, with Gini coefficient explaining over 60 % of the variation of proficiency in the long term and very considerably even in the very short and short terms.

Second, we note the stronger positive association between equality of opportunity and the level of outcomes on longer time frames. It implies an interaction across age groups between equality of learning opportunity and level of learning outcomes. This where we turn next, as interaction across age groups may arise due to reduced measurement error, in which case it should not have a dominant direction over time. If the interaction of equality of learning opportunity and proficiency across age groups has a dominant direction, it would suggest that both are linked to some common causal process, with either equality of learning opportunity or proficiency causally explaining the other or both reacting to a third cause with differing time delay.

The third observation – closely related to the previous one – is that Gini is a good predictor of relative learning outcomes in the short and very short terms even for the high parental education type. It should be noted that here Gini is not cohort-specific, but rather a medium-term indicator in our framework. The predictive ability of Gini implies an interaction between longer-term equality indicators and shorter-term learning outcomes.

The next section ***10.2*** presents the results of the analyses of the direction of the association between equality of opportunity and the level of learning outcomes across age groups and time, before ***Chapter 11*** draws some strands of research together and explores models for predicting proficiency with equality indicators.

10.2 Direction – Equality predicts proficiency across time

After having established that equality of learning opportunity and proficiency are positively associated, and observing that the association is stronger over longer time frames, it is time to assess the direction of the association between equality of learning opportunity and proficiency in time.

The five analyses show that earlier levels of equality of learning opportunity are positively associated with later levels of proficiency of the high parental education type. The analyses between age groups and within data sets seem to indicate that equality of learning opportunity and proficiency have a mutually predictive relationship, in which each predicts the other, but equality of learning opportunity is more predictive of proficiency than vice versa. The analyses across data sets find that the association of equality of opportunity and the proficiency of the high parental education type goes both ways.

Predicting level by level within a data set

Level of equality in older age groups is positively associated with the level of proficiency in younger high parental education age groups. The 0-year difference condition, with a slope of 0.08 and coefficient of determination of 0.017, shows the same – very weak – positive association between equality of learning opportunity and proficiency in the very short term that was observed earlier in Table 2.1a, when proficiency was predicted by equality of learning opportunity in the same age group and data set. The association of equality of learning opportunity across age groups leads us to expect that the association would change when going from contemporary prediction within the age group and data set.

It can be observed that the explanatory power of proficiency increases slightly when equality of learning opportunity in the next age group is predicted, but decreases when the distance to the predicted age group grows. It becomes statistically insignificant when the gap grows to 40 years, where the equality of learning of the 16-24-year-olds is predicted by the proficiency of the 55-65-year-olds. This demonstrates that the stronger positive association of proficiency and equality of learning opportunity over longer terms does not arise due to the level of proficiency predicting equality of learning opportunity in younger age groups. (***Table 2.3a***)

For the prediction in the other direction, the picture is somewhat different. Proficiency is more strongly associated with equality in older age groups than in the same age group, with both a steeper

Table 2.3a: Prediction across age groups

		Equality predicts proficiency				Proficiency predicts equality			
predicted age group	**n**	**slope**	**SE**	**R^2**	**p**	**slope**	**SE**	**R^2**	**p**
x	710	0.08	0.03	0.008	0.017 *	0.10	0.04	0.008	0.017 *
x+10	568	0.22	0.04	0.064	0.000 ***	0.17	0.04	0.024	0.000 ***
x+20	426	0.29	0.04	0.089	0.000 ***	0.15	0.05	0.020	0.004 **
x+30	284	0.27	0.05	0.079	0.000 ***	0.10	0.06	0.008	0.131 -
x+40	142	0.31	0.08	0.102	0.000 ***	0.06	0.08	0.004	0.450 -

slope of association and a higher coefficient of determination. The steepness of the slope rises to about threefold and the coefficient of determination 10-12-fold from the baseline of contemporary prediction. The difference in the strength of the association between proficiency predicting equality of learning opportunity and equality of learning opportunity predicting proficiency is especially apparent when the age group difference is 20 or more years.

The analysis shows that, when observed across age groups within data sets, the positive association of equality of learning opportunity and proficiency has a dominant direction, with equality of learning opportunity being a more powerful predictor of proficiency across age groups than vice versa.

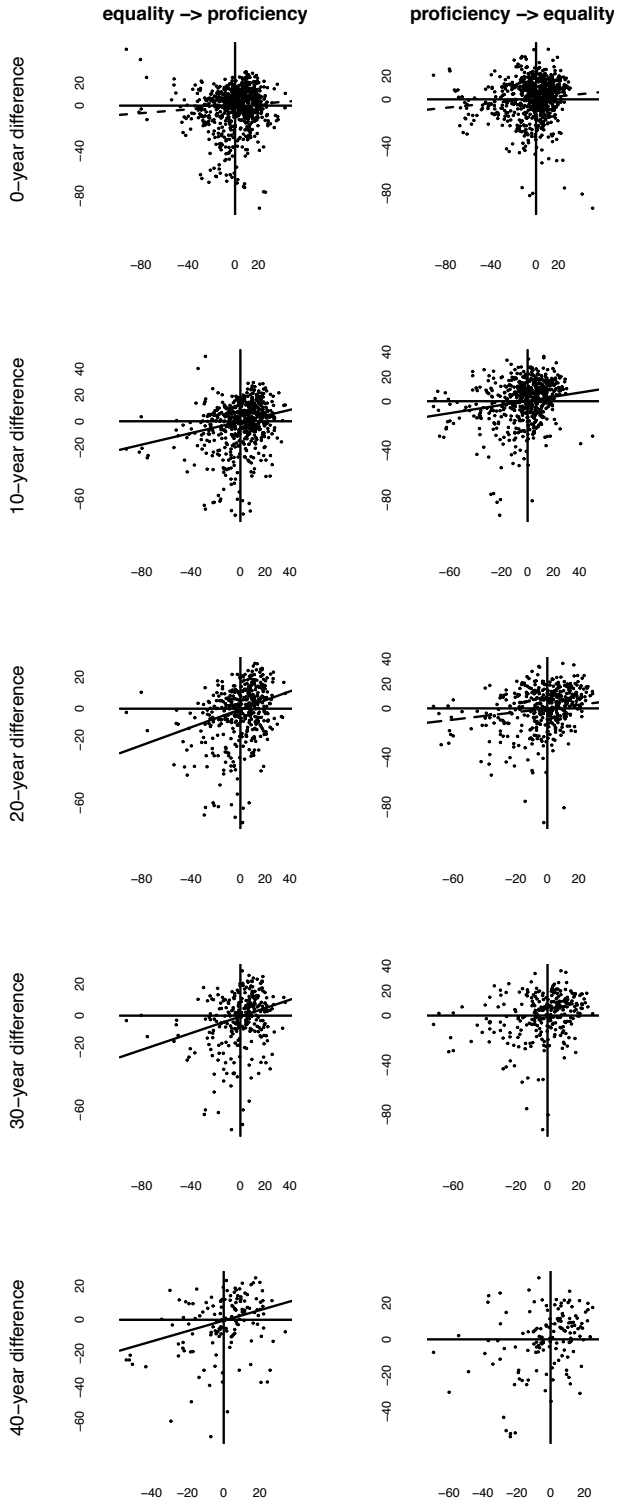


Figure 2.3a (on the right): Equality of learning opportunity explains proficiency across age groups.
x-axis=dependent & y-axis=independent variable;
regression line: ***=solid, **=long dashes, *=dashes

Predicting change by change within the data set

The next analysis predicts changes in equality of learning opportunity using changes in proficiency of the high parental education type and vice versa. The results (***Table 2.3b, Figure 2.3b***) show clear differences between different time frames in this association.

For simultaneous change (0-year difference) from old age group to younger ones (equality of learning opportunity difference between age groups a and b are compared to proficiency differences between the same age groups), changes in equality of learning opportunity are negatively associated with changes in proficiency. For all changes of 10, 20 or 30 years, more equality in the younger age group is also associated with a simultaneous change towards lower proficiency of the high parental education type. This would imply an equality-quality trade-off for that type in the short run.

The picture changes when the change between older age groups is used to explain the change between younger age groups. Many of the associations are statistically insignificant, but overall the negative association of changes of equality of learning opportunity to changes in proficiency turns into a positive association, whereby improvements in equality of learning opportunity predict later improvements in proficiency. Improvements in the proficiency of the high parental education type are, for the most part, associated with simultaneous decreases in equality of opportunity. They also predict later improvements in equality of opportunity, but the predictions are mostly statistically insignificant.

Again, as with levels predicting levels, changes in equality of learning opportunity are a better predictor of changes in proficiency between younger age groups than vice versa. For a 10-year change, the difference is smaller than for changes over the longer run. For a 20-year change, the picture is somewhat clearer, with equality of learning opportunity as the stronger predictor for 10-year difference and the considerably stronger predictor for the 20-year difference. For 30-year difference, the lagged results are statistically insignificant in both directions.

This analysis also suggests that, while equality of opportunity and proficiency form a positive loop, the positive long-term association between equality of learning opportunity and proficiency does not primarily arise due to improvement/decline in proficiency of the high parental education type predicting subsequent improvement/decline of equality. Furthermore, it suggests that, even if equality of learning opportunity is positively associated with the proficiency of the high parental education type in the long term, in the short term changes the group faces a trade-off where an increase of equality of learning opportunity seems to be associated with a simultaneous decline in proficiency. Interestingly, this short-term trade-off would seem to pay off in the longer run, as improvements in equality of learning opportunity seem to be associated with improvements in proficiency further down the line even for the high parental education type.

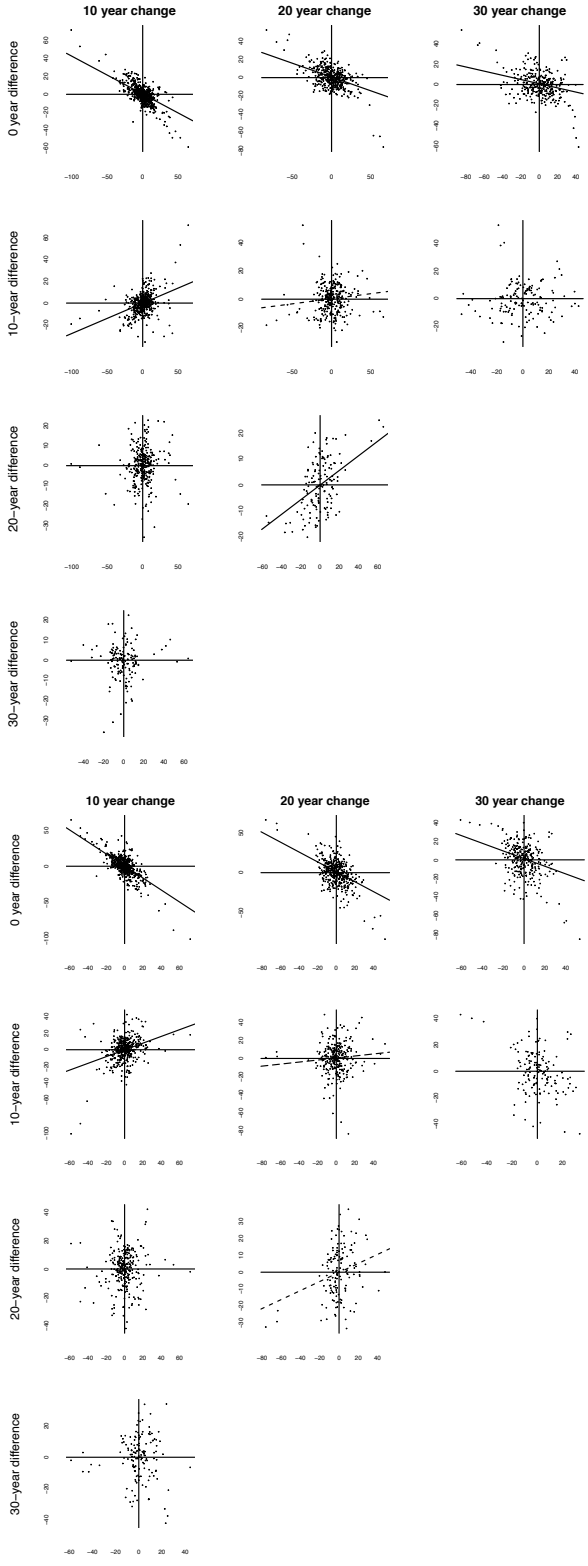


Figure 2.3b (on the right): Change of equality of learning opportunity explains change in proficiency across age groups

equality -> proficiency above, proficiency -> equality below

x-axis=dependent & y-axis=independent variable
regression line: ***=solid, **=long dashes, *=dashes

Table 2.3b: Analysis 1B: Change of equality of learning opportunity between older age groups explains the change in proficiency between younger age groups

age group difference	10 years change					20 years change					30 years change				
	n	slope	SE	R^2	p	n	slope	SE	R^2	p	n	slope	SE	R^2	p
Equality of opportunity predicts proficiency															
0	568	-0.42	0.02	0.350	0.000 ***	426	-0.30	0.03	0.187	0.000 ***	284	-0.20	0.04	0.084	0.000 ***
10	426	0.29	0.03	0.149	0.000 ***	284	0.08	0.04	0.015	0.038 *	142	0.08	0.06	0.013	0.171 -
20	284	0.03	0.04	0.002	0.471 -	142	0.30	0.06	0.154	0.000 ***					
30	142	-0.01	0.06	0.000	0.845 -										
Proficiency predicts equality of opportunity															
0	568	-0.83	0.05	0.350	0.000 ***	426	-0.62	0.06	0.187	0.000 ***	284	-0.41	0.08	0.084	0.000 ***
10	426	0.40	0.06	0.088	0.000 ***	284	0.11	0.09	0.006	0.191 -	142	0.14	0.12	0.010	0.239 -
20	284	0.07	0.07	0.004	0.303 -	142	0.27	0.09	0.053	0.006 **					
30	142	-0.10	0.11	0.006	0.350 -										

Note:

* = $p < 0.05$

** = $p < 0.01$

*** = $p < 0.001$

Predicting PIAAC level by IALS level

Using IALS levels of equality of learning opportunity /proficiency to predict PIAAC proficiency/equality of learning opportunity gives somewhat different results than across age-group analysis within data sets. (***Table 2.3c, Figure 2.3c***)

IALS equality of opportunity and proficiency predict PIAAC proficiency and equality of opportunity, respectively, with statistical significance on both jurisdiction and age-group levels. An increase in equality of learning opportunity by narrowing of the achievement gap by one point in IALS is associated with an increase in proficiency of 0.27 (jurisdiction) or 0.20 (age group) points in PIAAC. Equality of learning opportunity in IALS explains 19.7 % (jurisdiction) and 14.3 % (age group) of the variation of proficiency in PIAAC. The opposite association is somewhat stronger on the jurisdiction level, with IALS proficiency of the high parental education type explaining some 20.7 % of the variation of equality of opportunity in PIAAC. On the age group level, IALS equality of opportunity is a much better predictor of PIAAC proficiency than vice versa.

Removing the outliers Chile and the USA weakens the observed association over time. While the slopes of the association remain mostly unchanged on the jurisdiction level, the associations become statistically insignificant.

On the age group level, statistical significance is not lost, but the impact of equality of opportunity on later proficiency is significantly reduced so that the strength of the prediction is mostly equivalent in both directions.

These results suggest that there is a mutually reinforcing relationship between equality of opportunity and the level of learning outcomes, with the relative level of one predicting the relative level of the other over time, but also that equality of opportunity is a somewhat better predictor of future proficiency than vice versa. However, the impact of Chile and USA on the results indicates that a broader set of jurisdictions would be needed for more robust conclusions to be drawn.

Table 2.3c: Equality of learning opportunity level in IALS predicts proficiency level in PIAAC

		equality predicts proficiency				proficiency predicts equality			
	n	**slope**	**SE**	**R2**	**p**	**slope**	**SE**	**R2**	**p**
Jurisdictions in IALS and PIAAC									
jurisdiction	36	0.27	0.09	0.197	0.007 **	0.35	0.12	0.207	0.005 **
age group	180	0.22	0.04	0.143	0.000 ***	0.20	0.06	0.061	0.000 ***
Without Chile and USA									
jurisdiction	32	0.22	0.11	0.119	0.053 -	0.26	0.13	0.121	0.051 -
age group	160	0.15	0.05	0.061	0.002 **	0.22	0.07	0.060	0.002 **

Note:
* = p < 0.05
** = p < 0.01
*** = p < 0.001

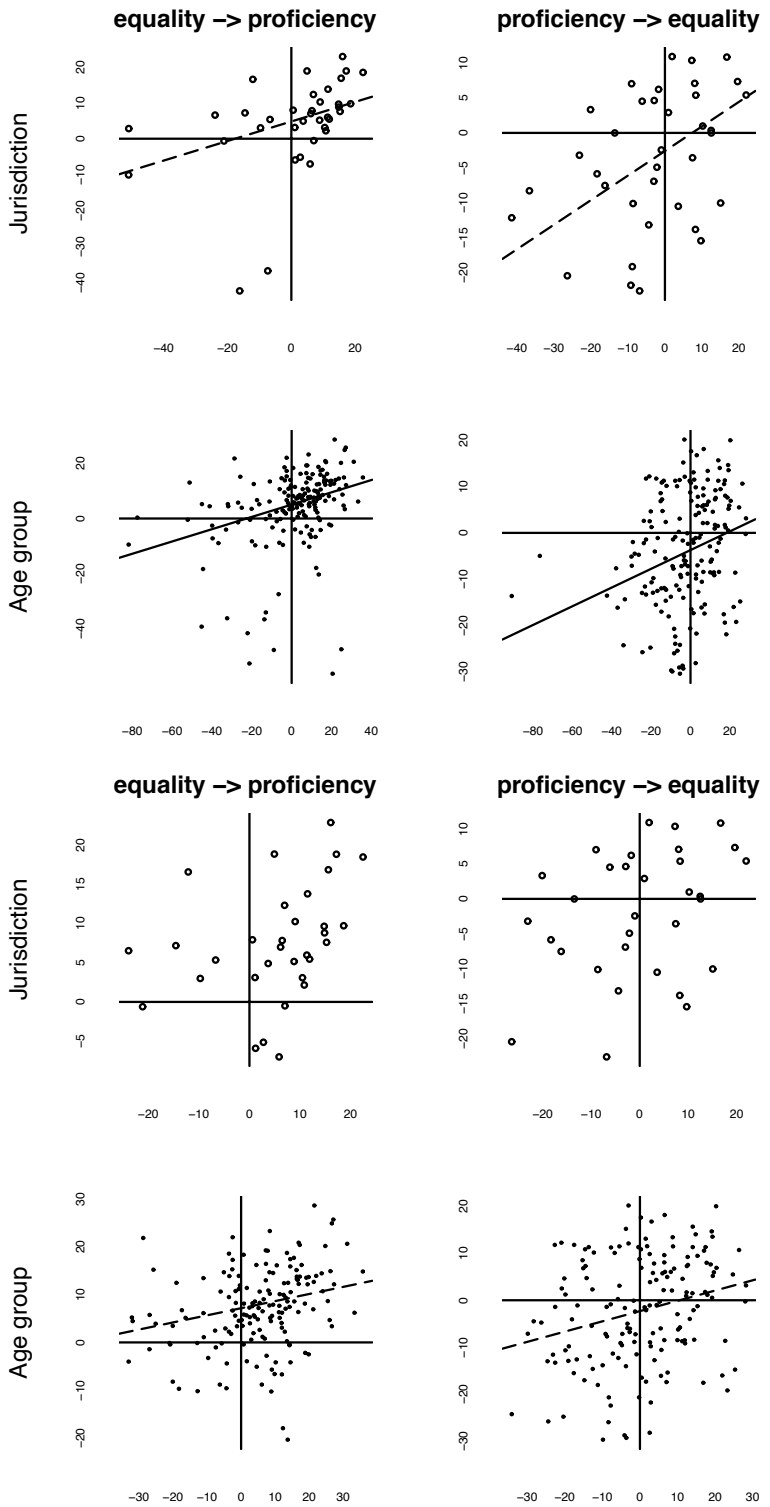


Figure 2.3c: Predicting PIAAC levels with IALS levels with (above) and without (below) Chile and United States.

x-axis=dependent & y-axis=independent variable

regression line: ***=solid, **=long dashes, *=dashes)

Predicting change by change across age groups and data sets

In the fourth analysis on the direction of the association between equality of learning opportunity and proficiency, change in younger age groups between IALS and PIAAC is predicted by a change between older age groups between IALS and PIAAC. (***Table 2.3d and Figure 2.3d***)

In simultaneous change, changes in equality of learning opportunity are negatively associated with changes in proficiency, but the association is not statistically significant. This changes when the change between older age groups is used to explain the change between younger age groups, as the positive association is most often statistically significant. The strength of the association depends somewhat on the age difference between the predicting and predicted age groups. As in some analyses above, the changes in proficiency are most strongly associated with changes in equality of learning opportunity of 20-year-older age groups. Overall, the differences are rather small, but equality of learning opportunity is consistently a better predictor of proficiency change between younger age groups than vice versa.

Chile and the USA, as outliers, have a significant effect on the results. Removing them from the analysis strengthens the ability of equality of opportunity change to predict change in proficiency slightly, but the more notable change occurs in prediction in the opposite direction. Removing outliers makes a change in proficiency of the high parental education type a good predictor of later change of equality of learning opportunity. This change is equivalent to the change in the previous analysis, in which PIAAC levels were predicted by IALS levels.

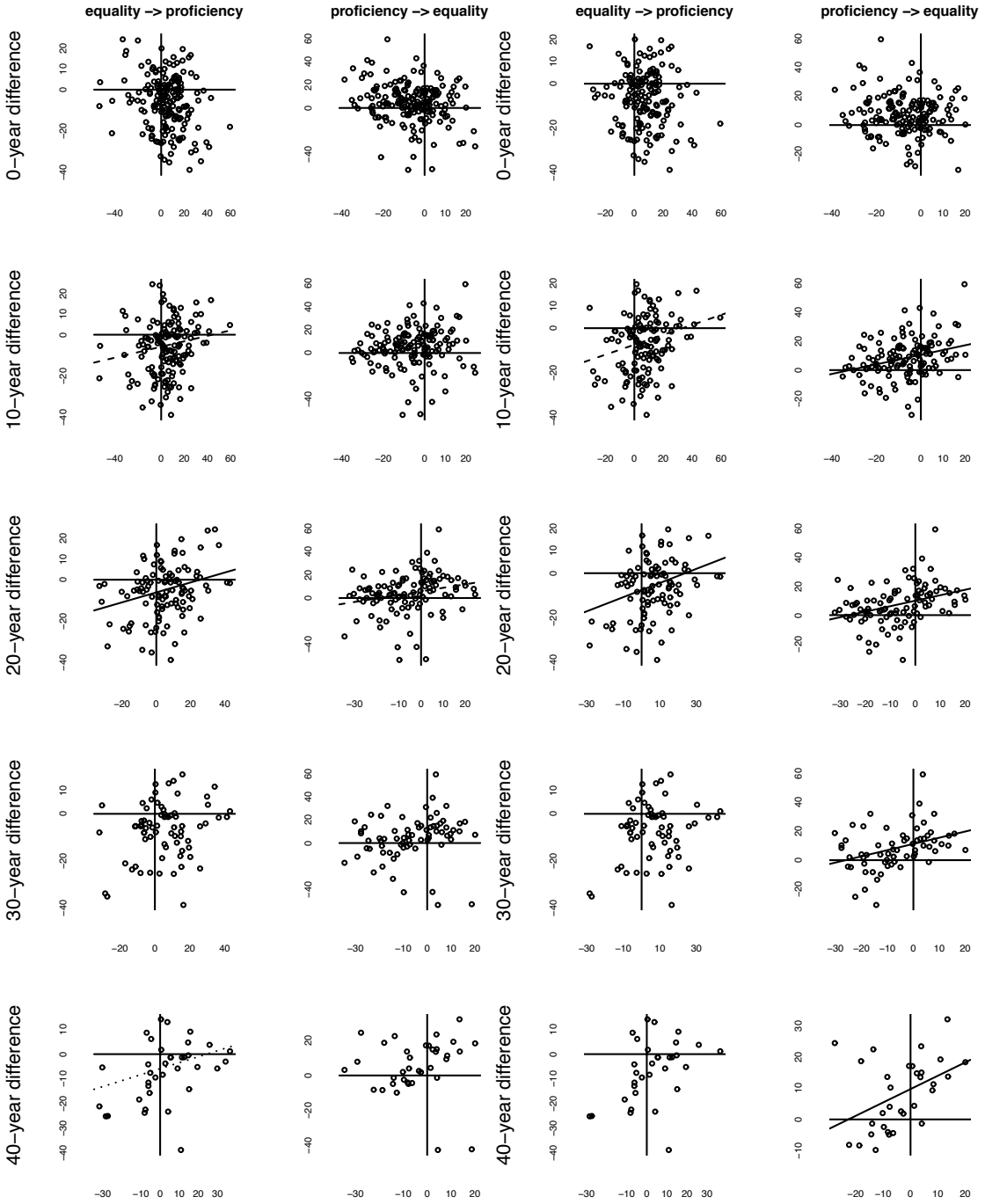


Figure 2.3d: Change between older age groups from IALS to PIAAC predicts change between younger age groups with (left) and without (right) Chile and United States.
x-axis=dependent & y-axis=independent variable; regression line: ***=solid, **=long dashes, *=dashes

Table 2.3d: Change between older age groups from IALS to PIAAC predicts change between younger age groups

difference between independent and dependent change (years)	n	Equality of opportunity predicting proficiency				Proficiency predicting equality of opportunity			
		slope	SE	R2	p	slope	SE	R2	p
Jurisdictions in IALS and PIAAC									
0	180	-0.06	0.05	0.007	0.256 -	-0.13	0.11	0.007	0.256 -
10	144	0.13	0.05	0.040	0.017 *	0.14	0.13	0.009	0.267 -
20	108	0.24	0.08	0.087	0.002 **	0.29	0.14	0.038	0.041 *
30	72	0.11	0.09	0.025	0.184 -	0.04	0.19	0.001	0.844 -
40	36	0.24	0.12	0.109	0.049 *	-0.19	0.24	0.019	0.426
Without Chile and USA									
0	160	-0.04	0.07	0.002	0.617 -	-0.04	0.09	0.002	0.617 -
10	128	0.22	0.08	0.063	0.004 **	0.30	0.09	0.082	0.001 **
20	96	0.32	0.10	0.105	0.001 **	0.36	0.09	0.135	0.000 ***
30	64	0.12	0.11	0.020	0.271 -	0.36	0.13	0.118	0.005 **
40	32	0.29	0.17	0.090	0.096 -	0.41	0.16	0.186	0.014 *

Note:

* = $p < 0.05$

** = $p < 0.01$

*** = $p < 0.001$

Overview of the analyses on the direction of the association

Our analyses provide evidence that equality of learning opportunity and proficiency of the high parental education type can act as predictors of each other over time and across age groups. Higher equality of learning opportunity predicts higher learning outcomes later on and in younger age groups, even for the high parental education type. The opposite also holds, and higher learning outcomes for the high parental education type predict higher equality of learning opportunity later on.

The results for the dominant direction of the association are inconclusive. Equality of opportunity is a better predictor of the level of proficiency in younger age groups and from IALS to PIAAC than vice versa. Also, the change of equality of opportunity between older age groups works better in predicting the change in proficiency between younger age groups than prediction in the opposite direction. Here the analysis is very sensitive to the jurisdictions in the analysis, as excluding Chile and the United States from the analyses reverses the dominant direction of the association, making proficiency change a better predictor of subsequent change in equality of opportunity, than vice versa.

Thus, the expectation that there is interaction across age groups between equality of learning opportunity and proficiency is confirmed, but whether there is a dominant direction to the association is much less sure. While, in most analyses, equality of opportunity is a better predictor of proficiency of the high parental education type than vice versa, the two seem to form a virtuous circle, and the available evidence does not allow us to conclude that one direction of the association is uniformly dominant over the other. It may yet be that the association does not have a dominant direction, and merely arises as greater precision is achieved with larger sample size.

What seems clear is that, for the high parental education type, we have interaction across age groups, which seems to imply an interplay of effects over different time frames. After all, equality of opportunity among older age groups is a better predictor of proficiency than equality of opportunity in the same age group. Moreover, improvements in equality of opportunity seem to be associated with subsequent deteriorations of learning outcomes, even when they are followed by later improvements. This suggests that equality indicators might be used to predict later levels of proficiency even for the high parental education type, for which the coincident short-term association between equality of opportunity and proficiency is the weakest.

11 Predicting proficiency with equality indicators

After establishing that equality of learning opportunity and proficiency are positively associated, and that, in this association, equality of learning opportunity and proficiency of the high parental education type interact across age groups, it is time to turn to the construction of a model to predict proficiency using equality indicators.

The previously presented results that 1) equality of learning opportunity is more strongly associated with proficiency over the long term; and that 2) the Gini coefficient, which is not age group-specific, is more strongly associated with proficiency in the short term than equality of learning opportunity, indicate that proficiency in individual age groups is linked to equality levels over the longer term. This implies that we can use longer-term indicators of equality to predict proficiency in individual age groups. This prediction proceeds in two phases. In the first phase, we predict IALS and PIAAC results using coincident indicators. In the second, proficiency is predicted over time, narrowing the population of cases to the 18 jurisdictions present in our IALS and PIAAC data.

11.1 Coincident models

First stage model

The first stage model that predicts proficiency using medium-term equality of learning opportunity explains proficiency in all parental education types with statistical significance. (***Table 2.4a***) As is to be expected, the prediction is better for lower parental education types than higher, and better for longer terms than for shorter terms. What may be noteworthy, and is consistent with earlier observations on the presence of several level effects in the interaction of equality of learning opportunity and proficiency of the high parental education type, is that medium-term equality of learning opportunity is a better predictor of short-term and very-short-term proficiency than short-term or very-short-term equality of opportunity (see tables 2.1a-d). The difference is most marked for the high parental education type, for which the association of proficiency with coincident shorter-term equality of opportunity is weakest and sometimes statistically insignificant.

Including the Gini coefficient improves the model, especially for the high parental education group, and the explanatory power of income inequality is quite similar for all parental education types in the very short to the medium terms. The improvement of the model, especially for the high parental education group entails that, while above only a very small share of the variation of proficiency of the high parental education type could be explained by short-term or very-short-term equality of opportunity, medium-term equality of opportunity and income inequality can explain a decent 27.5 % of the variation of very-short-term proficiency and 30.8 % of it in the short term.

It should also be noted that income inequality, when it is included in the model, catches much of the variation explained by equality of opportunity in the first-stage model. While equality of learning opportunity remains a significant and impactful predictor of proficiency for lower parental edu-

cation types, for the high parental education type equality of learning opportunity becomes statistically insignificant after inclusion of income inequality. We will later return to our interpretation as to why this should be.

For types with lower levels of education, medium-term indicators produce a smaller improvement of predictive power over the shorter-term indicators but represent an improvement, nonetheless.

Table 2.4a: First stage and baseline models for short (n=710) and very short (n=1420) terms

	low				middle				high			
	slope	**SE**	**R2**	**p**	**slope**	**SE**	**R2**	**p**	**slope**	**SE**	**R2**	**p**
First stage												
Short	1.39	0.06	0.469	0.000 ***	0.41	0.04	0.120	0.000 ***	0.25	0.04	0.054	0.000 ***
Very short	1.28	0.04	0.382	0.000 ***	0.35	0.03	0.082	0.000 ***	0.23	0.03	0.040	0.000 ***
Baseline												
Short term												
ELOp	1.01	0.05	0.338	0.000 ***	0.23	0.04	0.077	0.000 ***	0.01	0.04	0.027	0.718 -
Gini	-2.03	0.09	0.341	0.000 ***	-1.45	0.10	0.252	0.000 ***	-1.57	0.10	0.281	0.000 ***
Model			0.679	0.000 ***			0.329	0.000 ***			0.308	0.000 ***
Very short term												
ELOp	0.89	0.04	0.270	0.000 ***	0.18	0.03	0.050	0.000 ***	-0.01	0.02	0.020	0.666 -
Gini	-2.03	0.08	0.316	0.000 ***	-1.40	0.08	0.212	0.000 ***	-1.56	0.07	0.255	0.000 ***
Model			0.586	0.000 ***			0.262	0.000 ***			0.275	0.000 ***

Note:
* = p < 0.05
** = p < 0.01
*** = p < 0.001

Controlled model

The model is then tested by constructing a control model that includes variables for the natural logarithm of GDP per capita and the social rank score. We also include a running variable for age group, with the value of the variable increasing with age.

Table 2.4b shows, first of all, that the predictive ability of the models is improved by the control indicators, but the indicators also catch some of the variation previously explained by equality indicators. This should be noted because the analysis does not attempt to show that any of our indicators act as well-defined causes of learning outcomes; it attempts to gauge whether higher learning outcomes co-occur with higher or lower levels of equality of opportunity or inequality in general. Thus, the analysis does not turn on equality of opportunity indicators remaining statistically significant to the inclusion of controls, which would be necessary if we tried to understand the mechanisms at play. We are interested in the movement of systems on the hypothetical Ability-Opportunity Curve, and even if that movement were driven entirely by the level of the GDP, our interest focuses on the shape of the curve and the position of current developed countries on that curve.

Secondly, the control models help us to illustrate the interplay of longer and shorter terms, as well as to link our results to previous literature. We may note that the level of GDP per capita is significant for all parental education groups and on all time frames, which matches the previous results of Nathkov and Kozina (2012) using PISA data.

Another point of interest are social rank scores, which capture the positional aspects of education and should allow us to observe whether the results are driven by changes in selectivity of tertiary education. If types are selected by innate ability, systems with more selective education systems should show offspring of tertiary graduates with correspondingly higher demonstrated ability. As can be seen, the rank scores are statistically significant in the short and very short terms, but their relative contribution to the results is very small. Interestingly enough, the social rank scores are negatively associated with proficiency, indicating that in a multivariate model, selectivity of higher education is inversely related to learning achievement. The pattern observed would also be expected if parental education represented an indicator of circumstances rather than innate ability. The negative association between the social rank scores of the high parental education group implies that more selective systems produce lower results, which would not hold if educational achievement were simply a reflection of innate ability. It would also be expected based on the fact that more developed countries often display higher levels of education and higher learning outcomes. The robustness of the effect to GDP indicator implies that the negative association is not entirely driven by differences in education levels linked to the level of economic development in general. But, as the effect of social rank is marginal, too much should not be made of it.

The observation of interest here is that, with this specification, medium-term equality of learning opportunity is statistically insignificant for the high parental education type. This would not seem to square with the earlier observation that short-term proficiency interacts with longer-term indicators, especially for the high parental education type. To observe what is going on, it is useful to construct one more model, including short-term equality of opportunity as one of the independent variables, before proceeding to prediction over time.

Table 2.4b: Controlled model for short (n=710) and very short (n=1420) terms

	low				middle				high			
	slope	**SE**	**R2**	**p**	**slope**	**SE**	**R2**	**p**	**slope**	**SE**	**R2**	**p**
Short term												
ELOp	0.89	0.04	0.284	0.000 ***	0.21	0.03	0.068	0.000 ***	0.05	0.03	0.024	0.166 -
Gini	-1.47	0.09	0.243	0.000 ***	-1.12	0.09	0.193	0.000 ***	-1.13	0.09	0.204	0.000 ***
GDP	13.90	0.89	0.217	0.000 ***	11.03	0.76	0.213	0.000 ***	9.96	0.82	0.199	0.000 ***
Age group	-1.24	0.37	0.002	0.000 ***	-0.58	0.33	0.001	0.000 ***	0.09	0.33	0.000	0.792 -
Social rank score	-0.53	2.10	0.025	0.001 ***	-5.42	1.53	0.047	0.081 -	-5.37	1.74	0.034	0.002 **
Model			0.771	0.000 ***			0.523	0.000 ***			0.461	0.000 ***
Very short term												
ELOp	0.78	0.03	0.224	0.000 ***	0.15	0.03	0.044	0.000 ***	0.02	0.03	0.017	0.495 -
Gini	-1.33	0.07	0.212	0.000 ***	-1.05	0.07	0.158	0.000 ***	-1.13	0.07	0.184	0.000 ***
GDP	15.14	0.72	0.229	0.000 ***	11.40	0.60	0.206	0.000 ***	9.75	0.63	0.182	0.000 ***
Age group	-0.69	0.15	0.003	0.000 ***	-0.31	0.13	0.001	0.016 *	0.04	0.13	0.000	0.762 -
Social rank score	-6.20	1.70	0.037	0.000 ***	-6.17	1.20	0.049	0.000 ***	-6.52	1.32	0.037	0.000 ***
Model			0.705	0.000 ***			0.458	0.000 ***			0.421	0.000 ***

Note:
* = p < 0.05
** = p < 0.01
*** = p < 0.001

Inclusion of the shorter-term equality of opportunity indicator changes the results somewhat (***Table 2.4c)***. For the low parental education type, both medium-term and short-/very-short-term equality of opportunity are positively associated with proficiency. Together, they account for a somewhat larger share of the variation of proficiency than medium-term equality of opportunity by itself. The absolute improvement is greatest for the low parental education type, but the relative improvement is most considerable for the high parental education type for which having two equality of opportunity indicators makes both statistically significant.

Including the shorter-term equality of opportunity indicator shows that while longer-term equality of opportunity is positively associated with proficiency of the high parental education type, shorter-term (short / very short) equality of opportunity is negatively associated with it. As we would expect from the finding that the positive association grows stronger in more extended time frames, the positive effect of the medium-term equality of learning opportunity is more substantial, though only slightly, than the negative effect of the shorter-term equality of opportunity. Thus, the statistical insignificance of equality of medium-term learning opportunity would seem to arise as the positive longer-term effect is masked by the shorter-term effect when only the longer-term indicator is included in the model. Or, to be more precise, while causes that lead to smaller achievement gaps in individual age groups can lead to lower learning outcomes, systematic causes that lead to lower gaps over the long term tend to lead to higher rather than lower results. This pattern persists irrespective of the phase in which short-term or very-short-term equality of opportunity is included in the model.

Table 2.4c: Controlled model with shorter-term equality of opportunity for short (n=710) and very short (n=1420) terms

	low				middle				high			
	slope	**SE**	**R2**	**p**	**slope**	**SE**	**R2**	**p**	**slope**	**SE**	**R2**	**p**
Short term												
ELOp medium	0.41	0.06	0.184	0.000 ***	0.20	0.06	0.050	0.001 ***	0.34	0.06	0.037	0.000 ***
ELOp short	0.48	0.04	0.182	0.000 ***	0.00	0.05	0.029	0.948 -	-0.29	0.05	0.019	0.000 ***
Gini	-1.44	0.08	0.216	0.000 ***	-1.12	0.09	0.185	0.000 ***	-1.09	0.09	0.201	0.000 ***
GDP	14.19	0.82	0.200	0.000 ***	11.02	0.76	0.210	0.000 ***	10.13	0.80	0.200	0.000 ***
Age group	-0.97	0.34	0.002	0.005 **	-0.58	0.33	0.001	0.082 -	0.16	0.33	0.000	0.630 -
Social rank score	0.45	1.93	0.024	0.817 -	-5.42	1.53	0.048	0.000 ***	-4.94	1.69	0.034	0.003 *
Model			0.807	0.000 ***			0.523	0.000 ***			0.490	0.000 ***
Very short term												
ELOp medium	0.39	0.04	0.155	0.000 ***	0.08	0.04	0.031	0.041 *	0.26	0.04	0.030	0.000 ***
ELOp very short	0.48	0.03	0.160	0.000 ***	0.09	0.03	0.023	0.002 **	-0.27	0.03	0.020	0.000 ***
Gini	-1.31	0.07	0.193	0.000 ***	-1.06	0.07	0.154	0.000 ***	-1.09	0.07	0.184	0.000 ***
GDP	15.76	0.65	0.218	0.000 ***	11.32	0.60	0.202	0.000 ***	9.95	0.60	0.183	0.000 ***
Age group	-0.47	0.13	0.002	0.002 **	-0.36	0.13	0.002	0.005 **	0.12	0.12	0.000	0.000 ***
Social rank score	-1.66	1.55	0.033	0.286 -	-6.13	1.19	0.049	0.000 ***	-6.78	1.28	0.037	0.311 -
Model			0.761	0.000 ***			0.462	0.000 ***			0.455	0.000 ***

Note:
* = p < 0.05
** = p < 0.01
*** = p < 0.001

Overview of coincident models

Equality indicators are robust predictors of proficiency in all coincident models, and a large share of the variation in proficiency between jurisdictions can be explained using indicators of equality of learning opportunity and income inequality. For the low parental education type, equality indicators are sufficient to explain over 65 % of the variation of proficiency in the short term and over 55 % in the very short term, while for the high parental education type, equality indicators explain 27-30 % of the variation of proficiency in the very short and short terms.

For different parental education types, the relevance of different equality indicators is somewhat different. For the high parental education type, income inequality has a much stronger relative effect on the predictive power of the model than it has for the low and middle parental education types. For all coincident models, medium-term equality of learning opportunity is positively associated with proficiency, while the very-short-term/short-term equality of learning opportunity has a more modulating role. For low parental education type, the very-short-term/short-term equality of learning opportunity is positively associated with proficiency, and for the high parental education type, the association is negative. For the middle parental education type, the very-short-term/short-term equality of learning opportunity is statistically insignificant.

These results are consistent with the previous results showing that equality of learning opportunity and proficiency are associated across age groups, and this association is stronger for groups with higher parental education. This association across cohorts is apparent not only in the significance of medium-term equality of learning opportunity in explaining the variation of proficiency over the very short and short terms but also in the significance of income inequality in explaining the variation of proficiency, especially for the high parental education type. It is important to remember that the indicator of income inequality is not age group or cohort-specific and is thus comparable to medium-term equality of learning opportunity in that it measures equality over a large number of cohorts.

After observing the relative importance of longer-term equality indicators in predicting proficiency over the very short and short terms, it is time to further elaborate our predictive model by including indicators that would predict proficiency not only across age groups but also over time.

11.2 Predicting proficiency across time

Analysis of the direction of the association between equality of learning opportunity and proficiency showed that equality of learning opportunity predicts proficiency not only across age groups but across data sets and time. The previous models showed that inclusion of a longer-term indicator of equality of learning opportunity and/or income inequality considerably improves the predictive power of the model. Both observations imply that the baseline model predicting proficiency using coincident equality of learning opportunity can be improved by adding indicators of earlier levels of equality of learning opportunity. In this phase, the analyses are limited to short-term cases.

Univariate prediction

Before proceeding to predictive models, it is useful to establish how well a univariate prediction of PIAAC proficiency by IALS or PIAAC equality indicators performs. First, it allows us to establish the performance of predicting PIAAC short term learning outcomes with PIAAC medium-term equality of opportunity as in the first stage model above, but only predicting PIAAC outcomes and for the set of jurisdiction present in both IALS and PIAAC. Additionally, it allows comparison of that univariate prediction to predicting PIAAC learning outcomes with IALS medium-term equality of learning opportunity, as well as predicting PIAAC outcomes with either PIAAC or IALS income inequality. The emerging pattern is familiar from what has been suggested by the analyses above.

The upper half of the upper left quadrant of ***Table 2.5a*** shows the performance of PIAAC medium-term equality of learning opportunity in predicting PIAAC short-term proficiency. Compared with predicting the learning outcomes in either IALS or PIAAC for the set of 710 short-term cases (***Table 2.4a***), the predictive performance is somewhat lower for the low parental education type and higher for both the middle and especially the high parental education types. Comparison with the initial results (***Table 2.1a***) predicting short term learning outcomes using short-term equality of opportunity shows the medium-term indicator doing considerably better for the high and middle parental education types and marginally better for the low parental education type.

The lower half of the same quadrant shows that, as earlier analyses suggested, income inequality does considerably better than equality of learning opportunity in predicting learning outcomes, predicting over 45 % of the between-jurisdiction variation for both high and middle parental education types and almost 60 % for the low type.

Of interest for the observation that longer-term equality indicators perform better than shorter-term ones in predicting proficiency are the results in the lower left quadrant of the Table 2.5a, predictions of PIAAC proficiency using IALS indicators. IALS equality of opportunity predicts PIAAC proficiency consistently better than PIAAC equality of opportunity. IALS income inequality outperforms PIAAC income inequality for the high parental education type, matches it for the middle type and is somewhat weaker for the low type. Both indicators suggest that proficiency is associated with longer-term equality in ways that make it possible to anticipate future proficiency using equality indicators.

For both coincident (PIAAC) and predictive (IALS) indicators, and for both predictors (equality of learning opportunity and Gini), Chile and USA are influential observations. Their removal from the analysis significantly flattens the slopes and reduces the coefficients of determination. Despite this, a statistically significant association remains between equality of opportunity and learning outcomes as well as a negative one between income inequality and learning outcomes. It may be noted as well that removal of Chile and USA actually improves prediction using equality of opportunity

indicators from both IALS and PIAAC for the low and middle parental education types, only reducing predictive power for the high parental education type. Income inequality is much more affected by the inclusion or exclusion of Chile and the United States, even if the negative association remains despite their removal from the sample.

The results predicting PIAAC results using IALS income inequality are somewhat sensitive to the choice of the year of the Gini coefficient. The sample of jurisdictions in both IALS and PIAAC includes four Nordic countries – Norway, Denmark, Sweden, Finland – in which income inequality grew significantly in the 1990s. Our method uses Ginis for the closest year available, which for all these countries falls after administration of the IALS and after the rise in income inequality. If the last data point preceding IALS data collection were used, the income inequality data would predate the rise of income inequality. Such a change improves the ability of the model to predict PIAAC proficiency very slightly and makes Gini a statistically significant predictor of proficiency for the middle parental type as well, even in the absence of Chile and the USA. Here we have used the Gini figures from the closest year to avoid overfitting of the model by data selection, and thus offer this small observation as input for future research.

Table 2.5a: Univariate prediction of PIAAC learning outcomes

	All (n=180)				Without Chile and USA (n=160)			
	slope	**SE**	**R2**	**p**	**slope**	**SE**	**R2**	**p**
PIAAC ELOp								
low	1.48	0.14	0.401	0.000 ***	1.02	0.09	0.422	0.000 ***
middle	0.63	0.11	0.164	0.000 ***	0.44	0.08	0.170	0.000 ***
high	0.40	0.09	0.108	0.000 ***	0.14	0.07	0.027	0.040 *
PIAAC Gini								
low	-2.73	0.14	0.596	0.000 ***	-1.39	0.27	0.145	0.000 ***
middle	-1.75	0.14	0.463	0.000 ***	-0.73	0.20	0.078	0.000 ***
high	-1.49	0.12	0.460	0.000 ***	-0.57	0.16	0.071	0.001 ***
IALS ELOp								
low	0.97	0.08	0.431	0.000 ***	0.86	0.07	0.484	0.000 ***
middle	0.40	0.05	0.232	0.000 ***	0.44	0.05	0.318	0.000 ***
high	0.27	0.04	0.171	0.000 ***	0.22	0.06	0.083	0.000 ***
IALS Gini								
low	-2.22	0.14	0.587	0.000 ***	-1.11	0.21	0.145	0.000 ***
middle	-1.48	0.12	0.464	0.000 ***	-0.50	0.17	0.054	0.048 -
high	-1.30	0.10	0.483	0.000 ***	-0.46	0.13	0.072	0.001 ***

Note:
* = p < 0.05
** = p < 0.01
*** = p < 0.001

Baseline and predictive baseline models

Next in turn are the baseline and the predictive baseline models, predicting PIAAC proficiency using either coincident (PIAAC) or earlier (IALS) equality of learning opportunity and income inequality. Before constructing an augmented model that would predict proficiency using both coincident and predictive indicators, we can hereby compare the baseline model to a predictive model that simply replaces contemporary indicators of short and medium-term equality of learning opportunity and income inequality with predictive indicators of the same. Thus, proficiency in PIAAC is predicted by short-term and medium-term equalities of opportunity and income inequality in IALS instead of the same indicators in PIAAC.

Table 2.5b shows that the bivariate models mostly improve prediction over the univariate models. For both coincident prediction and prediction over time, the improvement offered by the bivariate model over prediction by income inequality is small for the middle parental education type and non-existent for the high type. Improvements over univariate prediction using equality of opportunity are considerable. As before, the predictive model performs very much on the same level as the coincident one, but equality of learning opportunity performs somewhat better in prediction over time rather than coincidently, just as observed with univariate prediction. The success of the predictive model is notable because the data used predates the predicted data considerably. The predicting equality of learning opportunity data from IALS in 1994-1998 and the Gini index from the nearest possible year predates the predicted PIAAC data from 2012-2015 by 14-19 years. The fact that a quarter of the variation in proficiency of the high parental education type, a third of the variation in middle parental education type and over half of the variation of the low parental education type can be predicted by data that predates the predicted data by almost two decades would suggest that equality indicators hold promise for use in predicting later proficiency.

As before, removal of Chile and the United States cuts into the predictive power but prediction remains statistically significant. The predictive power of income inequality in particular is considerably reduced, by as much as 90 % in the case of middle parental education type in the predictive baseline model. This reduction in the effect of income inequality does not translate fully into lost predictive power for the model, as the statistical effect of inequality of opportunity on outcomes increases with the removal of Chile and the USA for both low and middle parental education types and in both models. For the high parental education type, removal of the outlier jurisdictions reduces the effect of both equality measures.

It is noteworthy that while bivariate models do not improve prediction when Chile and USA are included in the analyses, they do improve it when Chile and USA are excluded. The predictive baseline model outperforms univariate prediction when Chile and USA are excluded, especially for the high parental education type. The predictive ability is slightly better, but the improvement over univariate prediction is slightly lower if Ginis predating the rise of income inequality are used for the Nordic countries. This may suggest that the relevant structural factors proxied by Gini are even longer-term than measured by medium-term Gini coefficient.

We take this pattern to suggest that the two equality indicators tap into an underlying structure of inequality that is not fully captured by either indicator. Overall, the bivariate prediction performs rather well, with IALS equality indicators able to account for almost half of the variation of PIAAC proficiency.

Table 2.5b: Predictive and predictive baseline models

	All (n=180)				Without Chile and USA (n=160)			
	slope	**SE**	**R2**	**p**	**slope**	**SE**	**R2**	**p**
Baseline model								
low parental education								
PIAAC ELOp	0.93	0.10	0.269	0.000 ***	0.95	0.09	0.392	0.000 ***
PIAAC Gini	-2.20	0.15	0.464	0.000 ***	-1.08	0.21	0.115	0.000 ***
Model			0.733	0.000 ***			0.508	0.000 ***
middle parental education								
PIAAC ELOp	0.28	0.09	0.096	0.002 **	0.40	0.08	0.154	0.000 ***
PIAAC Gini	-1.58	0.15	0.395	0.000 ***	-0.56	0.19	0.061	0.003 **
Model			0.492	0.000 ***			0.215	0.000 ***
high parental education								
PIAAC ELOp	0.00	0.08	0.054	0.976 -	0.09	0.07	0.018	0.184 -
PIAAC Gini	-1.48	0.14	0.406	0.000 ***	-0.51	0.17	0.063	0.003 **
Model			0.460	0.000 ***			0.081	0.001 **
Predictive baseline model								
low parental education								
IALS ELOp	0.59	0.07	0.281	0.000 ***	0.80	0.07	0.440	0.000 ***
IALS Gini	-1.72	0.13	0.438	0.000 ***	-0.71	0.16	0.101	0.000 ***
Model			0.718	0.000 ***			0.541	0.000 ***
middle parental education								
IALS ELOp	0.17	0.05	0.133	0.000 ***	0.42	0.05	0.295	0.000 ***
IALS Gini	-1.57	0.13	0.366	0.000 ***	-0.20	0.15	0.031	0.168 -
Model			0.499	0.000 ***			0.326	0.000 ***
high parental education								
IALS ELOp	0.04	0.04	0.087	0.342 -	0.17	0.06	0.066	0.004 **
IALS Gini	-1.24	0.12	0.399	0.000 ***	-0.35	0.14	0.055	0.010 *
Model			0.486	0.000 ***			0.121	0.000 ***

Note:
* = p < 0.05
** = p < 0.01
*** = p < 0.001

Augmented model

The augmented model uses equality indicators from both IALS and PIAAC to predict PIAAC proficiency. It offers a very small improvement in predictive power over predictive baseline model. The improvement over the baseline model is slightly greater but still relatively modest. ***(Table 2.5c)***

Thus, the main contribution of the augmented model is not in the ability to outperform the predictive baseline model. It does outperform the predictive baseline model, but so barely that the improvement is hardly worth the amount of additional data required. Most notably, the additional data required is gathered at a much later time, which emphasises the strong performance of the predictive baseline model relative to models including coincident indicators.

However, using both coincident and predictive equality indicators in the same model allows us to observe that predictive indicators from IALS play a stronger role in predicting proficiency of the high and middle parental types – where they contribute some 2/3 of the predictive power of the model when Chile and USA are excluded – than for the low parental education type, for which predictive indicators contribute some ½ of the predictive power of the model.

We also observe that the relative importance of Gini for the predictive power of the model increases when we progress from the low towards the high parental education type and that in all cases both coincident and predictive Gini tend to have very similar weights.

As above, the dropping of Chile and the USA from the analysis reduces the predictive power of the model considerably, even if the models remain statistically significant for all parental education types. While IALS equality of opportunity mops up some of the predictive ability of IALS Gini and becomes statistically significant even for the high parental education type, the reduction in statistical power is very significant as the predictive ability is reduced from almost 49 % to 12.5 % of the variation of outcomes.

The sensitivity of that prediction to inclusion or exclusion of Chile and the USA suggests that it is necessary to better understand to what degree their inclusion or exclusion reveals an association that is hidden without them, as the variation of equality of opportunity and income inequality is smaller in their absence, and to what degree it produces spurious correlations.

Table 2.5c: Augmented model

	All jurisdictions (n=180)				Without Chile and United States (n=160)			
	slope	**SE**	**R2**	**p**	**slope**	**SE**	**R2**	**p**
low parental education								
PIAAC ELOp	0.68	0.13	0.155	0.000 ***	0.51	0.11	0.223	0.000 ***
PIAAC Gini	0.04	0.69	0.223	0.959 -	-0.23	0.6	0.053	0.707 -
IALS ELOp	0.27	0.09	0.148	0.003 **	0.51	0.09	0.264	0.000 ***
IALS Gini	-1.73	0.55	0.228	0.002 **	-0.56	0.48	0.055	0.242 -
Model			0.755	0.000 ***			0.596	0.000 ***
middle parental education								
PIAAC ELOp	0.25	0.11	0.057	0.022 *	0.14	0.08	0.086	0.073 -
PIAAC Gini	-0.06	0.58	0.185	0.920	-1.29	0.47	0.042	0.007 **
IALS ELOp	0.09	0.06	0.075	0.137 -	0.38	0.06	0.227	0.000 ***
IALS Gini	-1.22	0.47	0.197	0.011 *	0.80	0.39	0.024	0.043 *
Model			0.515	0.000 ***			0.379	0.000 ***
high parental education								
PIAAC ELOp	0.06	0.10	0.028	0.511 -	-0.02	0.08	0.010	0.788 -
PIAAC Gini	0.04	0.52	0.192	0.945 -	-0.41	0.48	0.031	0.395 -
IALS ELOp	0.02	0.06	0.049	0.752 -	0.19	0.07	0.054	0.011 *
IALS Gini	-1.27	0.42	0.218	0.003 **	-0.04	0.39	0.030	0.919 -
Model			0.487	0.000 ***			0.125	0.000 ***

Note:
* = $p < 0.05$
** = $p < 0.01$
*** = $p < 0.001$

Controlled model

In the controlled model we add the same controls as before: GDP per capita and social rank score. This improves the predictive ability of the model, but only very slightly. This also entails that the control indicators - mainly GDP per capita but in some cases also social rank score – take on some of the variation explained by equality indicators in the simpler models.

Just as before, removal of Chile and the USA from the analyses reduces predictive power, but the main equality indicators remain statistically significant by and large, even after the addition of numerous control variables and removal of the outlier jurisdictions. After removing Chile and the United States, medium-term equality of opportunity in IALS explains 9.5 % of the variation of proficiency between jurisdictions in PIAAC, suggesting that measured equality of opportunity is not, even with these specifications, in a trade-off with the level of proficiency in developed countries.

Equality of learning opportunity and income inequality as predictors of proficiency

Overall, the positive association of proficiency with equality of opportunity and its negative association with income inequality are robust to the inclusion of controls. Overall equality of learning opportunity is positively (and income inequality negatively) associated with proficiency for all parental education groups. In all analyses, the positive association of equality with the level of learning outcomes is more apparent for longer-term indicators. Medium-term equality of opportunity outperforms very-short-term and short-term equality of opportunity in predicting proficiency even in the very short and short terms. In the same vein, indicators of earlier equality of opportunity and income inequality outperform coincident ones.

The observed (y-axis) and predicted (x-axis) values of baseline, predictive baseline and augmented models are presented in ***Figure 2.4a***. (Piñeiro, Perelman, Guerschman, & Paruelo, 2008) The roughness of the model needed to predict a considerable share of the variation of outcomes is apparent in the figure, as the minimal number of predictive indicators produces considerable banding of the cases. First, as medium-term indicators predict short-term outcomes, one indicator of medium-term equality of opportunity predicts proficiency in 5 cases of 10-year age bands. As all the domains share the same Gini coefficient, one Gini coefficient is used to predict proficiency in 10 (PIAAC) or 15 (IALS) separate cases. The banding is pronounced, especially in the baseline and predictive baseline conditions that both depend on only two indicators. It is somewhat mitigated in the augmented model, where the same two indicators are used, but both from both IALS and PIAAC.

When several versions (medium and short term, predictive and coincident) of the same indicators (equality of learning opportunity, Gini) are used in the analyses, all versions of the equality of learning opportunity are not usually positively associated with proficiency, nor are the indicators of income inequality all negatively associated with proficiency. Instead, some versions of the indicator dominate the results, resulting in the overall positive association of equality of learning opportunity with proficiency and negative association of income inequality with proficiency, while other versions of the same indicator modulate the effect of the dominating indicator. This same pattern was already observed with the coincident model, where medium-term equality of learning opportunity established a positive association between equality of learning opportunity and proficiency, but short-term or very-short-term equality of learning opportunity modulated the effects of the longer-term indicator.

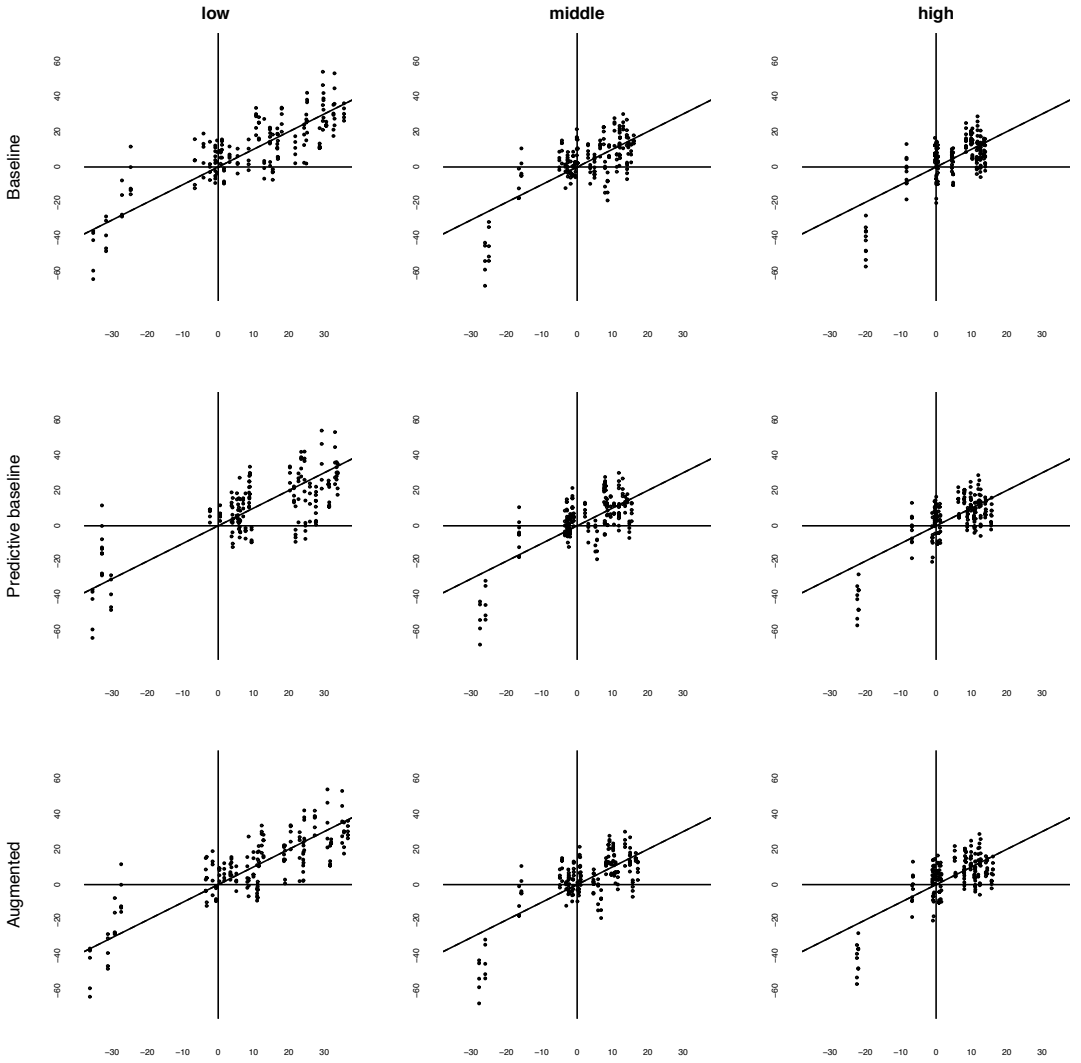
For income inequality, whenever only predictive (IALS) or coincident (PIAAC) indicators are used in the analyses, income inequality is uniformly negatively associated with proficiency. For ex-

ample, when both predictive (IALS) and coincident (PIAAC) indicators are used, the predictive indicator is negatively and the coincident indicator positively associated with the proficiency of the high parental education type, with the predictive indicator dominating in the model.

Table 2.5d – Controlled model

	All jurisdictions (n=180)				Without Chile and United States (n=170)			
	slope	SE	R2	p	slope	SE	R2	p
low parental education								
PIAAC ELOp	0.55	0.21	0.112	0.011 *	0.17	0.18	0.161	0.328 -
PIAAC Gini	0.42	0.85	0.175	0.622 -	0.98	0.72	0.037	0.174 -
PIAAC GDP	53.56	23.17	0.094	0.022 *	-74.56	19.82	0.087	0.000 ***
PIAAC Age group	-1.36	0.60	0.004	0.025 *	-0.73	0.51	0.003	0.154 -
PIAAC SRS	-1.20	6.68	0.010	0.857 -	-9.22	5.43	0.013	0.092 -
IALS ELOp	0.23	0.11	0.112	0.044 *	0.74	0.11	0.210	0.000 ***
IALS Gini	-1.97	0.70	0.181	0.005 **	-0.94	0.57	0.039	0.104 -
IALS GDP	-30.47	17.95	0.073	0.091 -	-65.66	15.24	0.067	0.000 ***
IALS SRS	8.01	6.36	0.020	0.209 -	-21.10	6.48	0.039	0.001 ***
Model			0.782	0.000 ***			0.655	0.000 ***
middle parental education								
PIAAC ELOp	-0.48	0.17	0.043	0.005 **	-0.26	0.13	0.056	0.049 *
PIAAC Gini	1.77	0.88	0.159	0.046 *	0.69	0.69	0.025	0.322 -
PIAAC GDP	155.04	29.07	0.069	0.000 ***	79.34	23.10	0.047	0.001 ***
PIAAC Age group	-0.67	0.50	0.004	0.183 -	-0.74	0.40	0.012	0.068 -
PIAAC SRS	-28.72	8.10	0.017	0.001 ***	14.30	6.33	0.012	0.025 *
IALS ELOp	0.41	0.10	0.075	0.000 ***	0.58	0.08	0.192	0.000 ***
IALS Gini	-2.57	0.72	0.164	0.000 ***	-0.32	0.58	0.017	0.579 -
IALS GDP	-110.64	22.69	0.055	0.000 ***	-63.48	17.71	0.042	0.000 ***
IALS SRS	-13.72	4.24	0.029	0.001 **	-17.24	3.66	0.075	0.000 ***
Model			0.616	0.000 ***			0.481	0.000 ***
high parental education								
PIAAC ELOp	-0.35	0.12	0.026	0.004 **	-0.24	0.10	0.015	0.013 *
PIAAC Gini	0.86	0.59	0.174	0.147 -	0.04	0.53	0.027	0.941 -
PIAAC GDP	121.18	28.14	0.027	0.000 ***	46.06	25.49	0.017	0.073 -
PIAAC Age group	0.35	0.48	0.001	0.468 -	0.19	0.40	0.001	0.641 -
PIAAC SRS	35.75	7.91	0.019	0.000 ***	18.65	6.70	0.019	0.006 **
IALS ELOp	0.27	0.07	0.056	0.000 ***	0.34	0.07	0.079	0.000 ***
IALS Gini	-1.75	0.48	0.194	0.000 ***	-0.21	0.43	0.026	0.471 -
IALS GDP	-93.60	21.60	0.025	0.000 ***	-44.67	18.75	0.019	0.018 *
IALS SRS	-30.96	5.59	0.048	0.000 ***	-23.76	4.88	0.101	0.000 ***
Model			0.571	0.000 ***			0.304	0.000 ***

Figure 2.4a: Proficiency Observed and predicted by the full model (x-axis=predicted, y-axis=observed; regression = black line)



Across all models, coincident equality indicators are relatively more important in predicting proficiency in the low parental education type and predictive indicators perform relatively better in the high and middle parental education types. Thus, even if the augmented model outperforms baseline and predictive models for all parental education types, the improvements over the second-best models (baseline for low parental education type and predictive for high and middle parental education types) are very limited. Improvements relative to the weaker models (predictive for low parental education type and baseline for high and middle parental education types) are somewhat more considerable. If data for the augmented model were not available, baseline model would seem to be a very good second-best for the low parental education type, and the predictive model for the high and middle parental education types. Even when data for the augmented model is available, the modest advantage

that the augmented model provides over the predictive baseline would speak for using the predictive baseline model.

We take these results to imply that even developed countries still sit on the upward-facing slope of the Ability-Opportunity Curve and thus do not face a trade-off between equality of opportunity and the level of learning outcomes. The robustness of this interpretation depends somewhat on how we should include Chile and/or the USA in the analyses.

Chile and the United States

Inclusion or exclusion of Chile and the USA in the analyses has a considerable effect on the results. While income inequality and equality of opportunity can be used to predict learning outcomes even in their absence from the analyses, the predictive power is considerably curtailed.

There are two relevant perspectives at play. As far as we are aware, neither Chile nor the USA are outliers due to faulty observation, inaccurate recording of the observation or by being affected by some influential unobserved factor, such as the subjects being drunk while taking the survey. As such clear causes for removing a faulty observation are absent, the aim would be to construct a model that should be able to include and explain Chile and the USA as well, even though they lie quite a distance from the main pack of countries.

On the other hand, removal of significant outliers is often prudent, and when it comes to the performance of the high parental education type relative to the level of equality of opportunity, both Chile and the USA are clearly outliers. (See ***Figures 2.1 a-d***) The USA is an outlier in some of the analyses as it has very high proficiency, especially in the high parental education type, despite having low equality of opportunity. Chile – in contrast – has much lower proficiency than the level of observed equality of opportunity would lead us to expect from the general association between the two.

Neither jurisdiction is much of an outlier when it comes to the low and middle parental education types. They are mostly outliers when it comes to the high parental education type. The fact that both become outliers mainly for higher parental education types suggest that the model is, to some degree, ready to cope with both when it comes to the low parental education type.

The presence or absence of the USA and Chile affects considerably the extent of the variation of equality of opportunity and income inequality between jurisdictions. Thus, a reduction in the predictive power of income inequality may arise from the fact that with less variation in income inequality it likely has less effect on the variation of outcomes.

In Figure 2.4a, Chile lies far from the main group of jurisdictions on both the predicted and observed levels of outcomes. The USA is closer to the main pack, but not far from being an outlier in the absence of Chile. ***Figure 2.4b*** illustrates the results when we predict proficiency for all 18 jurisdictions with a model constructed only with the 16 jurisdictions after excluding Chile and the United States. In many respects, the results are very similar, but the model constructed without Chile and the USA predicts overly high results for Chile, especially when we come to the middle and high parental education types where the predicted proficiency is close to the y-axis while the observed proficiency is dozens of points below the axis.

Excluding Chile and the USA seems to predict a flatter slope for the association between equality indicators and proficiency because excluding Chile and the USA narrows the range of income inequality in the sample, which leads to the model predicting overly high proficiencies of high parental education type for countries with high income inequality. This would suggest that, rather than outliers causing spurious associations, Chile and the USA are cases that are helpful in revealing a

pattern of associations that is less clear when the variation of the independent variable, income inequality in this case, is smaller.

It is unsurprising that the model in ***2.4a*** is a better fit with the full data than the model in 2.4b. That is true by construction. But it can be argued that the results imply that exclusion of Chile and the USA makes the model weaker because it limits observed variation and thus makes it more difficult for the model to identify the extent of the association of proficiency with equality indicators correctly. Comparison of ***Figures 2.4a*** and ***2.4b*** shows that, while a model excluding Chile and the USA has a hard time correctly predicting the position of Chile, it does rather well in predicting the results for the USA.

When proficiency is predicted by the model excluding Chile and the USA together with IALS Gini coefficient, the predictive power of the model is very close to the predictive power of the full

Figure 2.4b: Proficiency observed and predicted by model excluding Chile and United States (x-axis=predicted, y-axis=observed; regression = black line, perfect match = dotted line)

model constructed with both the USA and Chile included in the sample. The results are presented in

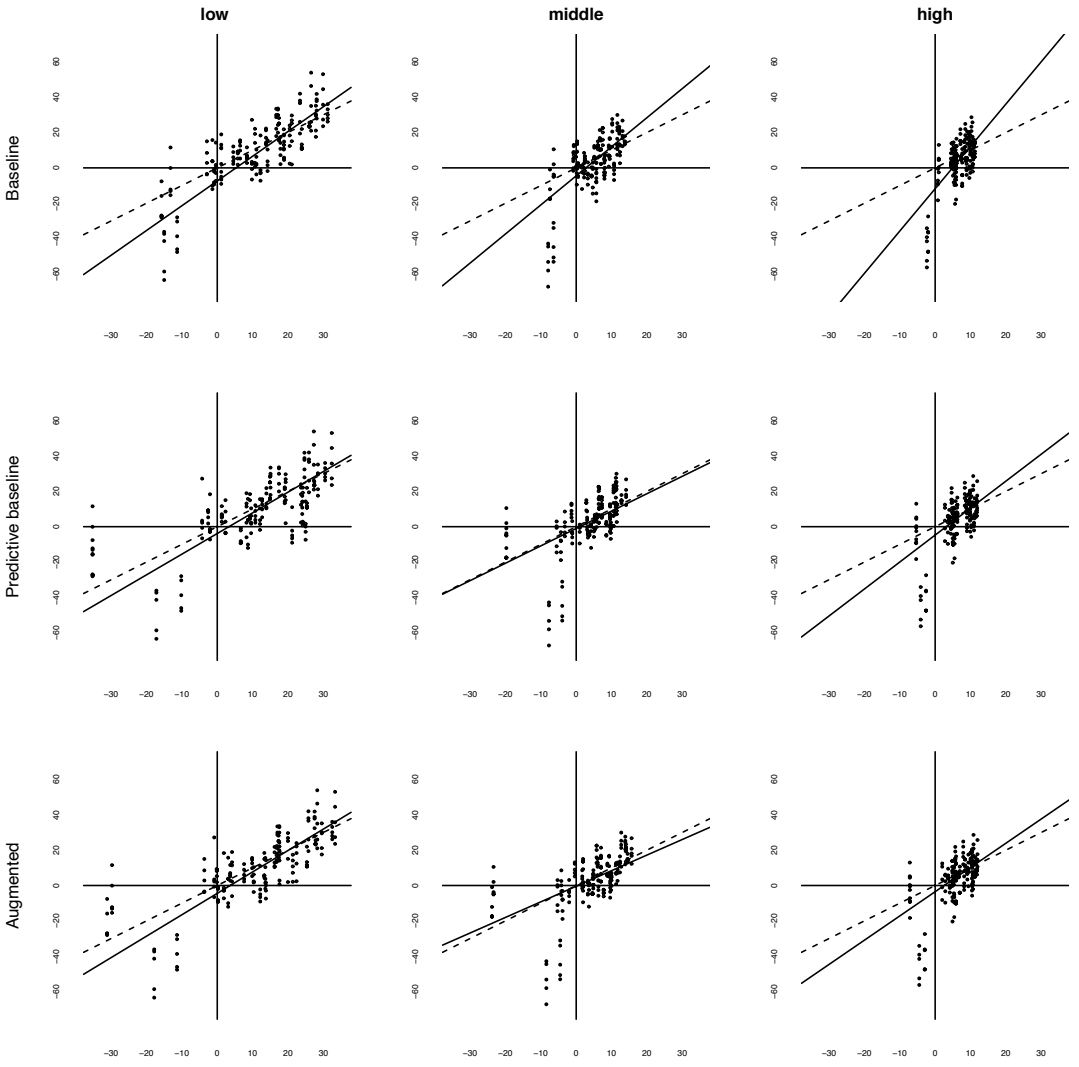


Table 2.5e and should be compared to ***Table 2.5c***, which presents the results of the augmented model. ***Table 2.5c*** shows the regressions for all 18 jurisdictions with a model made with all 18 jurisdictions and one predicting the outcomes of 16 jurisdictions without Chile and the USA using a model constructed without them.

Table 2.5e shows that the model constructed with 16 jurisdictions does quite well in predicting the outcomes for all 18 jurisdictions, as we already saw in ***Figure 2.4b***. The model allows us to predict some third of the variety of learning outcomes when Chile and the USA are in the sample but have not been used in the construction of the model. We take this to mean that a workable model can be identified even when the variation of equality of opportunity and income inequality are constrained by excluding extreme cases from the sample. But the inclusion of one indicator of income inequality considerably improves the fit of the model, almost reaching the same predictive power as achieved with the full augmented model and shown in ***Table 2.5c***. This is taken as an indication that it is indeed the effect of income inequality that is underdiagnosed if the USA and Chile are excluded from the model building.

Table 2.5e: Predicting proficiency with an excluding model and income inequality

	All jurisdictions (n=180)			
	slope	SE	R2	p
Excluding model				
low parental education	1.21	0.06	0.681	0.000 ***
middle parental education	0.88	0.10	0.306	0.000 ***
high parental education	-1.37	0.14	0.341	0.000 ***
Model with income inequality				
low parental education				
excluding model	0.83	0.08	0.421	0.000 ***
IALS Gini	-1.06	0.15	0.328	0.000 ***
Model			0.749	0.000 ***
middle parental education				
excluding model	0.39	0.10	0.219	0.000 ***
IALS Gini	-1.18	0.14	0.293	0.000 ***
Model			0.505	0.000 ***
high parental education				
excluding model	0.19	0.21	0.146	0.358 -
IALS Gini	-1.18	0.17	0.341	0.000 ***
Model			0.486	0.000 ***

We take these results as licence to interpret that the inclusion of Chile and USA in the sample used to construct a model does not produce spurious associations but rather allows better identification of associations of proficiency with equality indicators by allowing the observation of a much broader spectrum of equality of opportunity and income inequality.

Thus, we interpret that the analysis including Chile and the USA is to be preferred to analyses excluding Chile and the USA.

Overview

Analysis of models predicting proficiency over time shows that, when predicting PIAAC proficiency, coincident (PIAAC) indicators are superior to predictive (IALS) indicators in the low parental education type, but predictive indicators outperform coincident indicators in the medium and high parental education types. This result is consistent with the previously reported finding that longer-term

equality is a more significant predictor of proficiency in groups with higher parental education. In all the analyses and for all parental education types, income inequality explains about half of the variation explained by equality indicators.

The results of our analyses are sensitive to inclusion or exclusion of jurisdictions with notably low equality of opportunity or notably high income inequality. For the reasons presented above, we interpret that including such jurisdictions allows better recognition of the strength of the association between different aspects of equality and the level of learning outcomes.

Overall, the results consistently show that equality of opportunity is positively associated with the proficiency of the high parental education type. As much as half of the variation of proficiency of the high parental education type may be attributed to the variation in equality of opportunity and income inequality. This implies that, even in the relatively developed jurisdictions in our sample, we are still on the rising section of the ability-opportunity curve, where the observed socioeconomic differences in outcomes do not fully reflect between-type differences in innate ability. In these circumstances, improvements in equality of opportunity, or socioeconomic equality, should not be costly in the sense of being in a trade-off with the level of learning outcomes. Equality of opportunity and high quality can still be pursued – and achieved – at the same time.

12 Education policy and equality of learning opportunity

After showing that equality of learning opportunity and proficiency are associated, with equality of learning opportunity predicting proficiency over time and across cohorts, it is time to turn to the effects of education policy. The results of these analyses of the association of discontinuities in equality of learning opportunity with education policy reforms are presented in several phases. First, we present general results that set the path for a more detailed presentation. The presentation of the general results will be followed by a presentation of the results by jurisdiction, after which follows an overview of the results by type of education policy reform. Presentation of the general findings sets the scene, presentation of the results by jurisdiction takes a slightly more detailed look at each jurisdiction, and the last section draws together some of the common strands between the results.

12.1 General findings

Stability of equality of opportunity from IALS to PIAAC

In the first phase of the analysis, it is necessary to analyse to what degree the observed age group differences are cohort differences that can be used to link cross-sectional data on equality of learning opportunity observed at adulthood to longitudinal data on education policy. As suggested above, our approach is to see if the results produced by different domains, different calculation methodologies, different age groupings, as well as – most importantly – different surveys give consistent results on the evolution of equality of opportunity over time in different systems. To link IALS and PIAAC results to changes in education policy, the between-cohort differences in the level of equality of opportunity should remain relatively stable in adulthood over time.

To establish this, ***Figures 2.5a*** and ***2.5b*** display the level of equality of opportunity for 10- and 5-year age bands for both IALS and PIAAC, the former offering a more reliable measurement of the level of equality of opportunity in each case, the latter giving more granularity over time. The figures do not display equality of opportunity as the point difference to the international average, but the raw achievement gap. This choice is made to facilitate interpretation by the reader, but also because the comparison of raw achievement gap by age band in the two surveys produces essentially the same results as comparing gaps centred to the international age-group average.

In this phase, the pattern of discontinuities across age groups is inspected foremost to compare discontinuities obtained from IALS and PIAAC. As IALS and PIAAC cover some common age groups, their data has an overlap of 31-35 birth cohorts depending on the year of implementation of IALS. We can assess these cohorts with overlapping coverage for similarities in the pattern of discontinuities given by IALS and PIAAC domains. For each survey, the different domains give somewhat different results. While sometimes the results are very close to each other, in other instances, the difference between the two (PIAAC) or three (IALS) measures from the same survey can be non-negligible. Especially in the oldest cohorts, the three domains of the IALS produce somewhat differing achievement gaps, even compared to younger age groups in the same jurisdiction. This is good to

note when we compare the results for 10- and 5-year age bands as well as when we observe the progression of equality of opportunity over cohorts for the cohorts with data from both surveys.

We can observe jurisdictions where IALS and PIAAC give a notably different picture of the progress over time. Notable cases of this are Germany and the United States, for both of which IALS and PIAAC give differing accounts of the evolution of equality of opportunity in the cohorts born between the 1950s and 1970s. The difference is more prominent in the 5-year age bands. For Germany, the difference in 10-year age bands is relatively modest, and for the United States, the 10-year age bands give one data point in IALS that is not matched on a relatively same level by PIAAC. For our purpose of trying to link cohort differences to education policy, these two jurisdictions are fortunately rather unproblematic, as the federal structure – as well as the recent history of division into two social and political systems in the case of Germany – would pose significant challenges to linking any sudden cohort differences to changes in education policy in any case. But they offer a welcome reminder that our method relies on binned data with quite small samples in either the high or low parental education type, which translates directly into uncertainty in measuring equality of opportunity.

With the exception of these two notable exceptions, the progression of equality of opportunity is very similar in the IALS as it is in the light of PISA. Different jurisdictions will receive a more thorough presentation later on, but it may be noted already here that for Denmark, Finland and New Zealand, the two surveys paint an effectively identical picture of the progression of equality of opportunity across cohorts. Many other jurisdictions also show a high degree of conformity even in sequences of relatively small changes in the level of equality of opportunity. Notable examples of this are Italy and Ireland, both of which display signs of rises and decreases of equality of opportunity in cohorts born between the mid-1950s and the early 1980s. In some cases – as in Chile in the 5-year age bands – we can observe a very similar pattern of rises and declines, only slightly out of sync. Here it is useful to remember the binning of the data by the 5- and 10-year age bands of the surveys, which can produce asynchronies.

Overall, not only are the achievement gaps measured in the five domains of IALS and PIAAC consistently on the same level, the pattern of rises and declines is very similar. The relative consistency of the pattern of discontinuities indicates that cohort differences in the socioeconomic pattern of proficiency have remained relatively stable for the 14-21 years that passed for each jurisdiction between IALS and PIAAC. This would support the interpretation that the observed age group differences are, to a considerable degree, cohort differences. IALS and PIAAC show the same pattern of rises and falls across cohorts even when using data for 5-year age bands. This consistency suggests that 5-year age bands can be used to measure attainment gap, even though they have smaller sample sizes and are thus more susceptible to measurement error than the 10-year age band level.

Figure 2.5a: Achievement gap for 10-year age-bands by birth year (above)
Figure 2.5b: Achievement gap for 5-year age-bands by birth year (next page)


Belgium – Flanders
Chile
Czech republic
Denmark
Finland
Germany
Hungary
Ireland
Italy
Netherlands
New Zealand
Norway
Poland
Slovenia
Sweden
United Kingdom – England
United Kingdom – Northern Ireland
United States

Comparison with earlier research

One way to assess the robustness of the results on the achievement gap or the level of equality of opportunity across cohorts is to compare these results with earlier research measuring the same, or a closely related, construct with different data. Here we may turn to a variety of large-scale surveys of youth skills that have been conducted across developed countries over the last 50 years. Recent work by Anna Chmielewski (2019) has collected all the available international large-scale surveys since the early 1960s.

A general comparison of the results shows considerable convergence. The challenge is that the main results reported by Chmielewski concern the 90-10 achievement gap, not the gap between high-low parental education types used in our analysis. As noted above, the difference between the 90-10 and the high-low gaps is between taking parental education to be a positional or an absolute good. The difference in measurement may – and most likely does – produce some differences on the level of trends and their timing. Assessing the differences is impossible without more precise data, so here the comparison is on a very general level, noting any differences and not guessing which may be due to methodology and which to other causes.

Overall, the progression of the achievement gap across cohorts in ***Figures 2.5a*** and ***2.5b*** is similar to the results presented in Chmielewski (2019). For Flanders, the trend of a rising achievement gap from 1950 onwards, as demonstrated by Chmielewski, is more apparent in the 10-year age band data than the 5-year band data. For Chile, the rise and fall after 1960 turn into a new rise in our data but not in youth data. England shows a small initial decline and a later rise in both sources, as does Finland. The Czech Republic, Germany and Poland show rising gaps for youth, which match PIAAC but fit IALS more poorly or not at all. The same can be seen for Sweden, but the rise is much less pronounced when measured with adults. In part, this difference may arise as adult data covers cohorts not covered by youth data. The data from the Netherlands is difficult to interpret, but youth data shows a growing gap that is not apparent in the adult data. Denmark, Hungary, New Zealand and Norway show a rising gap in both data sources. Ireland shows a trend of rising gaps with youth data, but a smaller gap in younger cohorts with adult data. The United States shows a U-shaped curve with youth data, but the opposite with adults. Thus, we may note that there is broad consistency with earlier results, but that more precise interrogation of our results against previous research would necessitate comparison of our results with those of Chmielewski on a more detailed level, with a matching definition of the achievement gap.

For Finland, the comparison with Chmielewski's results is more accurate, as her results for Finland have been published (Salmela-Aro & Chmielewski, 2019) showing the results for each survey separately, rather than only the long-term trend as in the main article. (Chmielewski, 2019). Even more importantly, Salmela-Aro and Chmielewski (2019) report the results for the high-low parental education gap, rather than the 90-10 gap used in the main article. Their results match the U-shaped evolution of first declining and then rising inequality visible in the adult data. There is some discrepancy in timing, though, as the Chmielewski-Salmela-Aro results show a somewhat later rise in the achievement gap than the adult data does. In adult data, the rise affects cohorts born in the mid-1980s, while in youth data the first cohorts affected was born in the early 1990s. Note that the adult data seem to show somewhat earlier rises and falls than the youth data. This may arise partially due to the difference between the positional 90-10 gap and the high-low parental education gap, but it does not fully arise from positional measurement as it can be seen with the high-low gap for Finland as well.

A possible substantive cause for a difference in timing of trends is that changes in the education system may affect individuals through different levels of the education system. We are affected by

experiences during compulsory education, often in a comprehensive school, and through experiences after the cohort has been separated into different educational tracks. If the education system changes on several levels of education at once, we would expect it to start affecting different cohorts in the population in the education system through different levels of education. Those untouched by the changes in compulsory education would still be affected by the changes in upper secondary or in tertiary education. If such policy changes were to leave permanent traces, they would not be uniformly visible in youth and adult data. Youth data would show effects only after the cohorts affected through compulsory education were tested, but adult data would detect effects in older cohorts that were not affected by changes in compulsory education but were affected by upper secondary or tertiary. This would produce an appearance where the achievement gap measured somewhat later in life turns into decline earlier than the one measured earlier in life, if we observe the evolution of the gap cohort by cohort. While this is speculation, the mechanism is perhaps not implausible enough to warrant interpretation of this observation as an artefact.

When are the cohort differences established?

For changes in education policy to leave permanent marks in the form of cohort differences, it is also necessary that those differences become relatively fixed quite early. If they become relatively fixed early on, we may link cohort differences in outcomes to cohort differences in experiences over a relatively limited period of time. This is what is needed to link cohort differences to education policy reforms, even if the linking is only in terms of timing.

It may be observed in ***Figures 2.5a*** and ***2.5b*** that IALS and PIAAC results match quite nicely even for those cohorts that were the youngest at the time of IALS. For example, the youngest 5-year age bands (20-24 and 16-19) of IALS 1994 were born in 1975-1978 and 1970-1974 and were 38-42 and 34-37 at the time of PIAAC 2012. Despite the age difference and some differences in the measurement instruments, the achievement gap in PIAAC is virtually unchanged from IALS for most jurisdictions observed in both. This suggests that the achievement gap does not change significantly after quite early adulthood. As the youngest cohorts are just out of compulsory education, it is thus possible that age group differences can be taken as cohort differences.

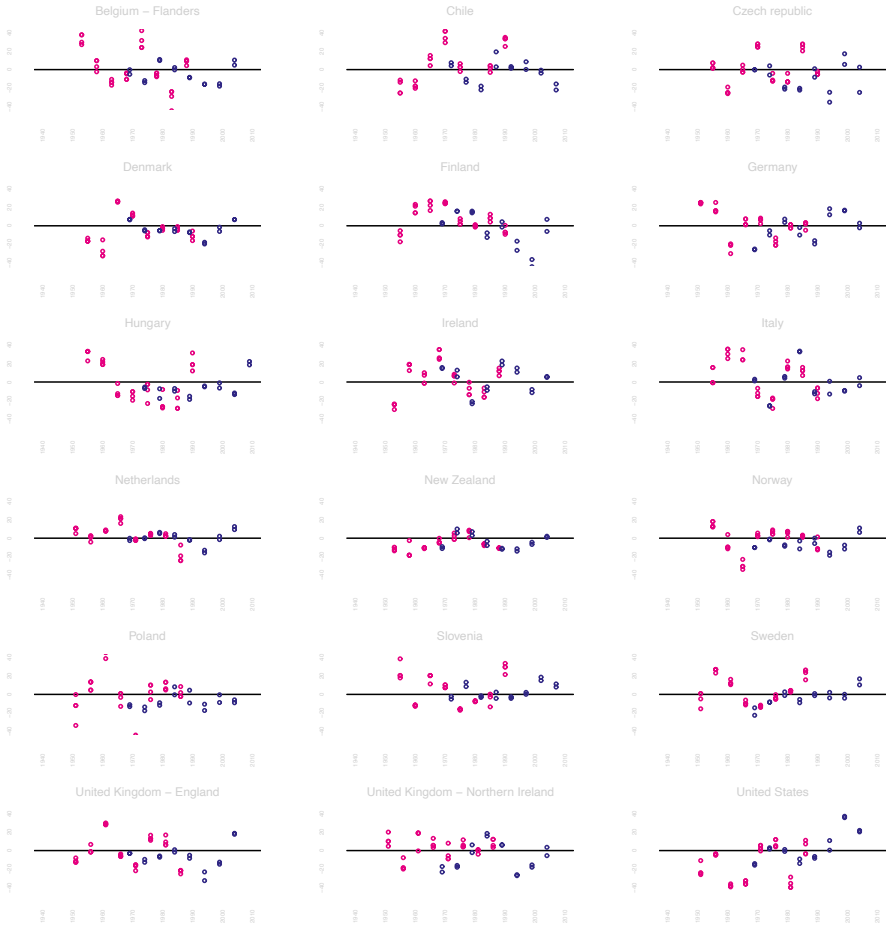
In addition to measurement error due to small samples, we may also note that there are some suggestions of age effects for the youngest age bands in some jurisdictions. E.g. in Sweden, the youngest 5-year age band displays a smaller achievement gap than the second-youngest in both IALS and PIAAC. This is suggestive of an age effect because the youngest 5-year age band of IALS is the only notable point where IALS and PIAAC data do not follow the same path. This may be due to age effects, i.e. due to the achievement gap growing in early adulthood in Sweden, when youth follow different tracks. The same pattern can be observed in Hungary. Such cases seem to be quite rare, which we take to suggest that by and large the level of the achievement gap is determined early enough to be observable in late teens or very early adulthood.

From levels to discontinuities

Now, to assess if changes in equality of opportunity can be linked to education policy reforms in different jurisdictions, it is useful to change the focus from levels in different cohorts to discontinuities between the cohorts. Here, instead of looking at levels, we focus on the change between two cohorts close to each other. As explained above in the section on methodology, our aim is to compare the last fully pre-reform cohorts to the first post-reform cohorts. This amounts to comparing cohorts born 10 years apart, as compulsory education has been some 10 years long in most countries, and thus the age difference of the cohorts with experience of only the pre- and post-reform systems should be on the order of 10 years.

The discontinuities are presented in ***Figure 2.5c***. The figure shows the discontinuities calculated with 5-year age bands. While the former prioritises accuracy of measurement for each cohort, the latter prioritises granularity over time and accepts greater noise as the price to pay. The interpretation of the results tries to take both aspects into account. Our general approach is that, if the sequence visible in 5-year discontinuities is not completely erratic, we tend to interpret that the pattern

Figure 2.5c: 5-year equality of opportunity discontinuities



is not an artefact produced by measurement and binning of the data. This inherently interpretive nature of the exercise is why we present the results by jurisdiction, even though the analysis does not attempt to be conclusive for any jurisdiction.

Sequence of improvement and deterioration

For the most part, the observed discontinuities form a continuous pattern over time. More substantial increases/decreases are preceded and/or followed by smaller increases/decreases for most jurisdictions. This pattern of progressively rising/diminishing discontinuities is consistent with the fact that there is overlap in the schooling experiences of the cohorts compared, and the effects of any reform are gradual. But in the 5-year discontinuities there are numerous examples where a sharp discontinuity of improved equality of learning opportunity is followed by first a stabilisation and then a discontinuity towards less equality of opportunity.

This pattern in which change is followed by an opposite change is sometimes very quick, as in Northern Ireland in the early 1950s, when an improvement is directly followed by a decline, but more often the change is more gradual, with the improvement followed by first stabilisation and then decline, as in Sweden after the mid-1950s, Ireland after the late 1960s, or Italy after the early 1960s. In many of these cases it is possible to identify a reform that matches, by its timing, the initial change in equality of learning opportunity, but more difficult to identify a reform that would fit the timing of the declining equality of learning opportunity. We hypothesise that this pattern could arise due to a single reform rather than a succession of several, balancing reforms if the effects of reform operate on slightly different time scales. If the low parental education types, due to lower levels of economic and cultural resources, are more responsive to any changes in education system than high parental education types, whose higher resources mitigate the effect of reforms, the pattern observed would arise due to the variation between types in their speed of response to the policy changes. The data used in this study do not allow us to assess this question. In general, we interpret such patterns as evidence of a single reform, rather than several consecutive ones.

12.2 Sequence by jurisdiction

Here we go through the sequence of discontinuities by jurisdiction, drawing attention very briefly to policy reforms that match, by timing, the discontinuities observed. While the results are more promising for some jurisdictions than others, we go through all the jurisdiction with the exception of the aforementioned federal systems of Germany and the United States, as any linking of discontinuities observed to policies pursued would be open to even more uncertainty than it necessarily is in any case.

As noted above, the timing of the discontinuities is based on the 5-year discontinuities, despite the inherently greater lability of the results arising from the smaller sample sizes for 5-year age bands than for 10-year ones. This raises the risk that the discontinuities observed arise from measurement error rather than actual differences between cohorts. Thus, we focus mainly on discontinuities that represent a significant rapid change in the level of the achievement gap or a change in the direction of the evolution of the gap.

The following pages present the size of the achievement gap by birth year, and the size of the discontinuity by the year of the implicit reform, both calculated as presented above in ***Chapter 9.2***. In all instances, we present the level of the achievement gap by birth year with the y-axis reversed, so that a rise represents improving equality of opportunity and a decline a deterioration in both the gap and the discontinuity figures. Below, we show the timing of the implicit reforms.

Belgium-Flanders may serve as an example of interpretation of the figures. ***Figure 2.6a*** shows that the cohort born in the early 1930s had a large achievement gap that was narrowed considerably by the cohort of the early 1940s (rise from the 1st to the 3rd point from the left). This shows up as a large positive discontinuity in the very early 1950s, as a reform in the very early 1950s would have left the cohort born in the earlyish 1930s untouched but affected the cohort born in the earlyish 1940s. Correspondingly, the fact that cohorts born after the mid-1980s show a larger achievement gap than slightly older ones shows up as negative discontinuities in the late 1980s and throughout the 1990s.

The association in time between observed discontinuities and education policy reforms mainly uses the two compendia of education reforms, and their data on the timing of reforms and the affected cohorts (Fort, 2006; Garrouste, 2010), as mentioned in ***Chapter 9***.

Belgium – Flanders

In Flanders, there is some discrepancy between IALS and PIAAC when it comes to 5-year discontinuities, which may affect interpretation.

In IALS we observe two large positive discontinuities, in the early 1950s and in the early 1970s, both times followed by stabilisation (change near zero) 5 years later and then decline 10 years later. The positive discontinuity in the 1970s is followed by a quick decline in the direction in the trend, while in PIAAC the turning of the trend from increasing equality of opportunity to increasing inequality is spread over a longer period.

The timing of the positive discontinuity in the early 1950s could match the timing of the 'school war' on public secular and Catholic education, and might be linked to the *Schulpakt* or *Pacte scolaire* of 1958 if it were to have had a very clear effect from the very first cohorts onwards.

The positive discontinuity in IALS in early 1970s would seem to correspond to the reform of 1971 towards a more comprehensive school system by postponing streaming to age 15/16 by introducing unique secondary schooling.

An optimist would also take the positive IALS discontinuity in the early 1980s and note that it corresponds nicely to the lengthening of the compulsory education to 18 years, but here the poor match between IALS and PIAAC can make us legitimately cautious.

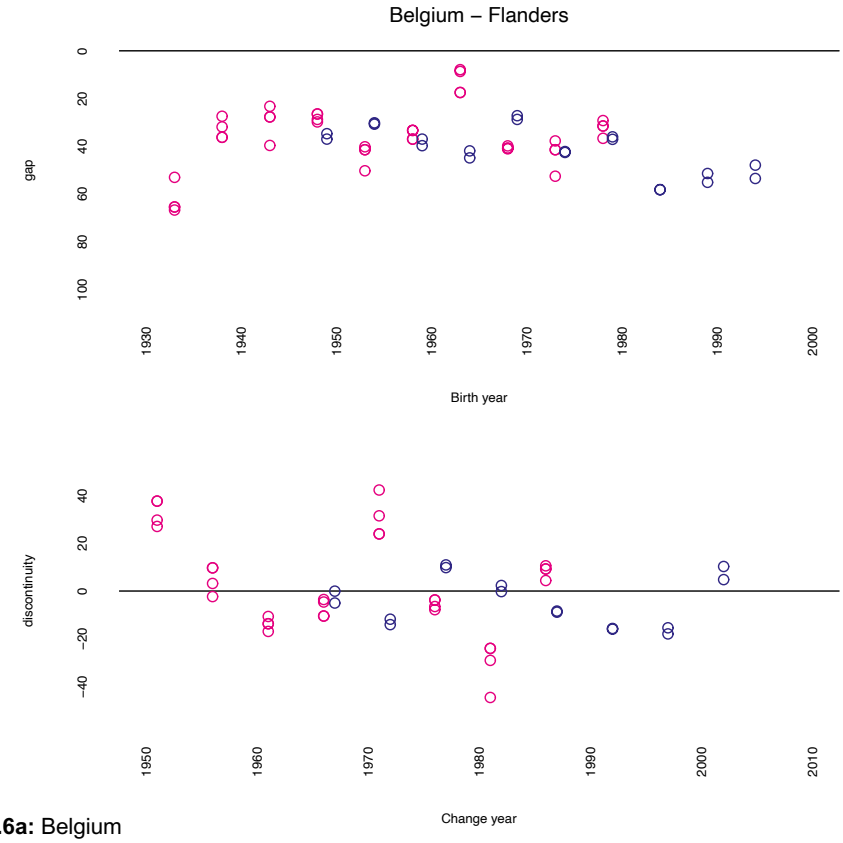


Figure 2.6a: Belgium
(level by birth cohort above, 5-year discontinuity below)

Chile

For Chile, the large discontinuities in the direction of improved equality of opportunity took place in the late 1960s and towards the late 1980s, both followed by a stabilisation and decline, especially in the case of the former discontinuity, where the late 1970s witnessed a very significant decline in equality of opportunity.

The former discontinuity coincides with the reforms of Eduardo Frei Montalva, during which primary education was practically universalised, and access to secondary and tertiary education increased. The latter would match the efforts in the very early 1990s to address questions of equity through numerous means, including increased funding for infrastructure and materials as well as targeted funding for rural areas and the poor. The position of teachers was also changed, and the change was accompanied by significant pay increases. (Delannoy, 2000)

It should be noted that the beginning of the 1990s saw substantial, more general changes with the democratisation of the country, so that linking the change to policies specific to education, even in general, must be done with caution.

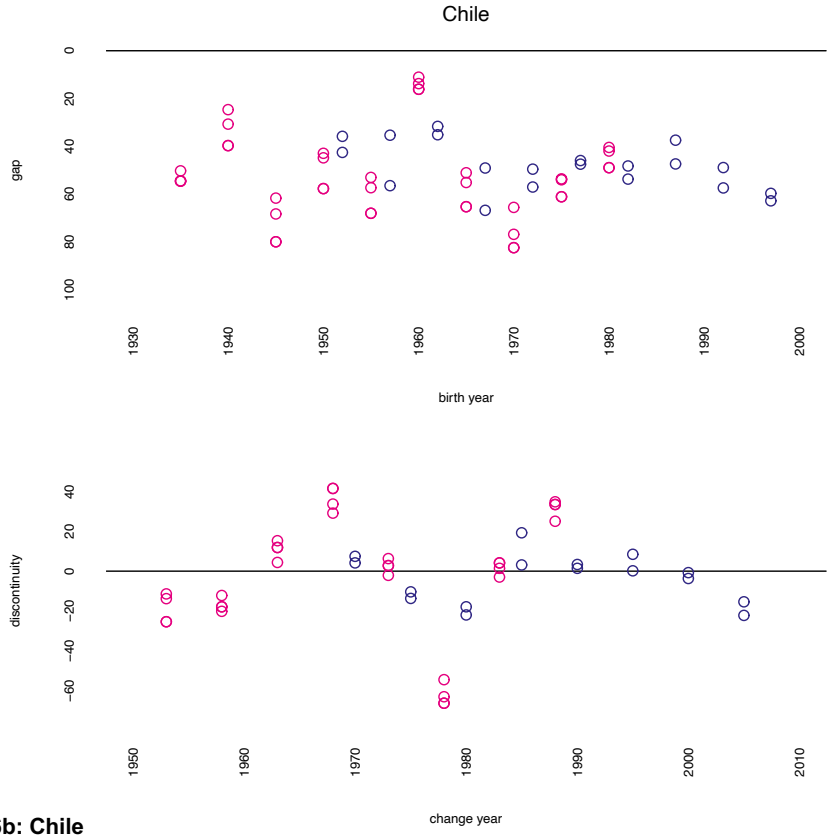


Figure 2.6b: Chile
(level by birth cohort above, 5-year discontinuity below)

Czech Republic

The development in the Czech Republic poses significant problems in linking discontinuities to specific reforms, as reforms in the length of compulsory education have been uncommonly frequent, especially before 1960. Over a relatively short period, compulsory education was thrice (1948, 1960, 1990) lengthened from eight to nine years and twice (1953, 1979) shortened from nine to eight years. With rough data, the frequent changes from 1948 to 1960 pose a problem for linking age group differences to specific reforms. The data resolution is simply insufficient to separate the effects of the reform of 1948, if they are visible, from the effects of the 1953 reform.

For later reforms, both IALS and PIAAC show a decline in equality of learning opportunity in the late 1970s or very early eighties, which corresponds in timing to the shortening of compulsory education in 1979.

PIAAC shows a decline in equality of opportunity in the early nineties, which may correspond to the lengthening of compulsory education in 1990. Here, as with some other post-communist countries, there is a notable discontinuity when the social and political system changes. This may be related to changes in education policy as well, but the sweeping social changes were so notable that one must be wary of overinterpreting the change.

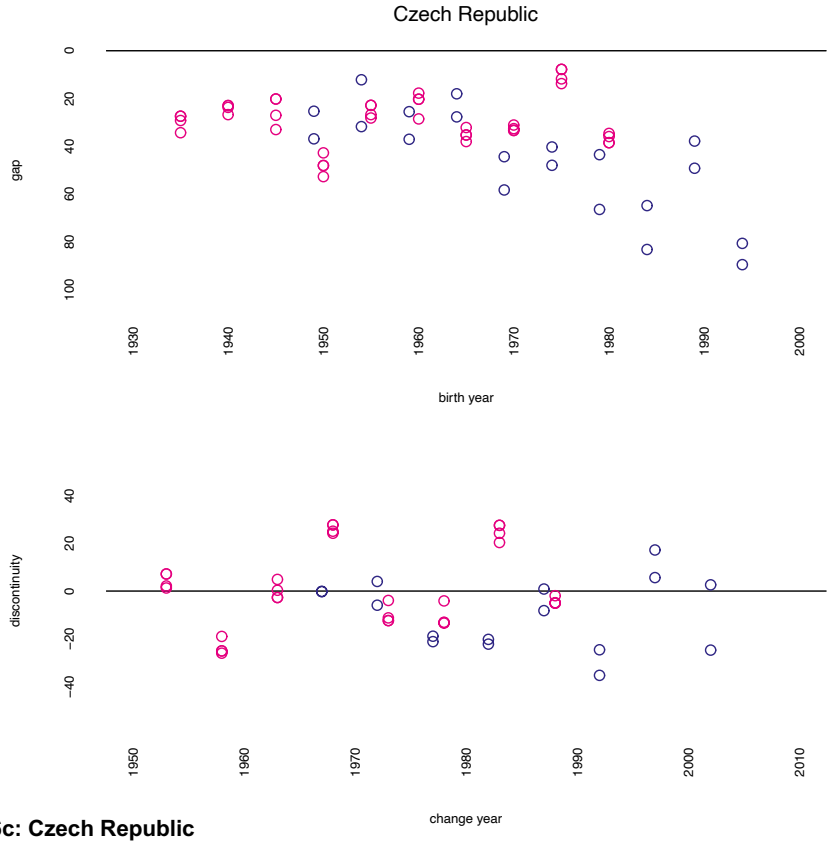


Figure 2.6c: Czech Republic
(level by birth cohort above, 5-year discontinuity below)

Denmark

In Denmark, IALS and PIAAC give a very similar, very stable picture of development for the cohorts for which they overlap. From mid-sixties to the early seventies we see improvements in equality of learning opportunity, followed by small reductions in the late seventies and a more significant decline in the late 1980s or early 1990s, as well as a small gain in the early 2000s.

The sharp rise in equality of learning opportunity in the early and late 1960s corresponds to the extension of compulsory education from seven to nine years in 1971 and would be consistent with the results of earlier research on the effects of extension in compulsory schooling as it captures the difference between cohorts that finished school in the early/late 1970s as opposed to the early/late 1960s.

Beginning in the early 1990s, we see a drop in equality of learning opportunity that corresponds – by timing – to the Danish reforms in 1993-1994. This would suggest that Denmark fits well among the Nordic countries in that it implemented education reforms in the 1990s that led to decreases in equality of learning opportunity. In Denmark, the change is less dramatic than in Finland, and more similar in size to the change experienced by Norway, also observed in the early 1990s. In seeking causes from specific education policy reforms, it must be noted that all the Nordic countries experienced rising income inequality in and after the early 1990s, and the analyses above suggest that income inequality is linked to both equality of opportunity and the level of learning outcomes.

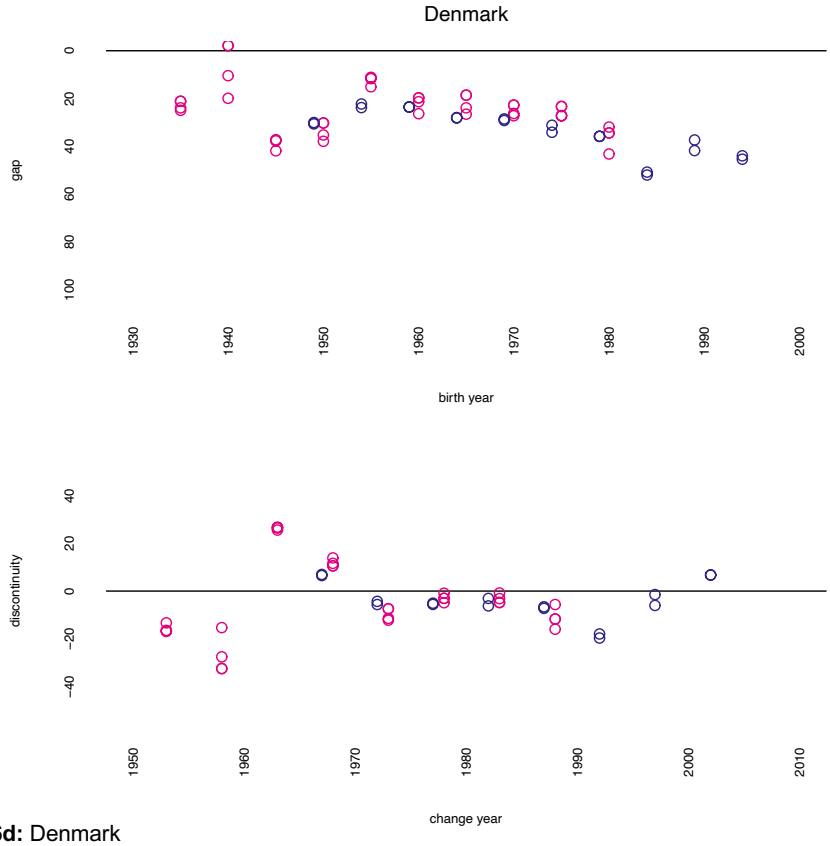


Figure 2.6d: Denmark
(level by birth cohort above, 5-year discontinuity below)

Finland

We will return to the case of Finland – as an example – in more detail below, but here we may note that we can observe improvement in equality of opportunity in the 1960s and in 1970s, the timing of which is a bit different in IALS and PIAAC. Moving IALS data points forward a few years would improve the match of the two surveys quite a lot in this case, but it is also essential to recognise that the underlying methodology is imprecise by nature, as it deals with rough binned data. The improvements since the late 1950s correspond, by timing, to the extension of compulsory education from 6 to 9 years in 1957, with the first affected cohort being born in 1951 and achieving the age of 16 in 1966. The improvements in the early 1970s match, in turn, the comprehensive school reform of the early 1970s, which has been shown to have increased proficiency and benefited the groups with lower socioeconomic background relatively more.

The early 1990s witness a decline in equality of opportunity - and the late 1990s a decline almost without parallel in the data. Notably, in comparison with the other Nordic countries, which display – as Denmark above – small changes even when changes occur, the decline in equality of opportunity from the late 1990s onwards is very significant. This is interesting, as many of the reforms of the 1990s have been, in popular accounts of the Finnish school system, credited for its success in many international large scale surveys, most notably PISA. We will return to the case of Finland later in some more detail, as an illustration of how the rough, indicative results that our methods produce may be used as clues for further research.

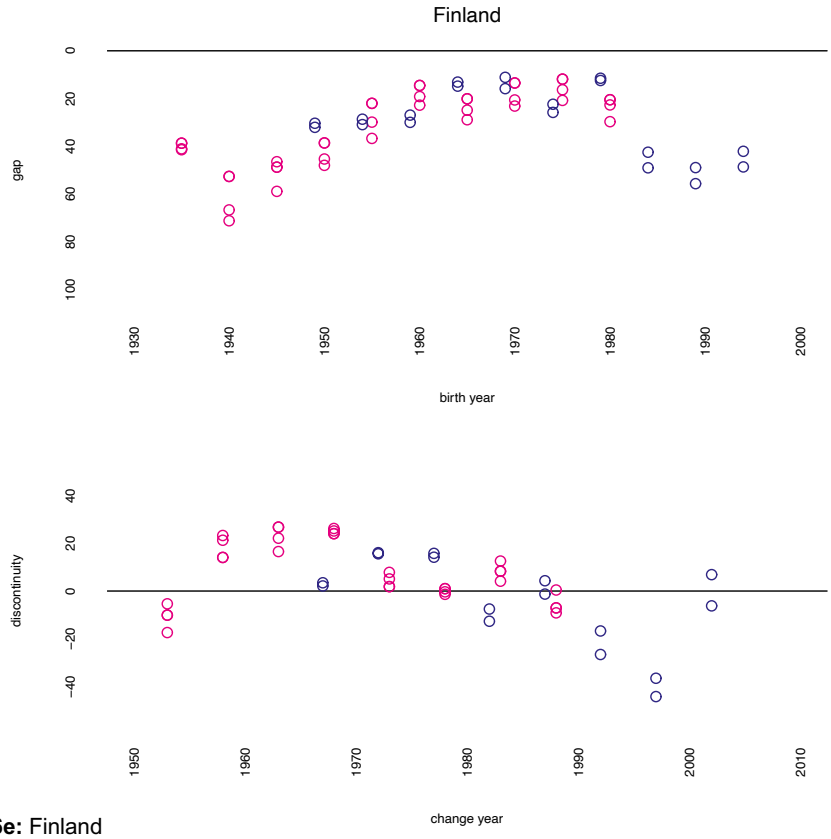
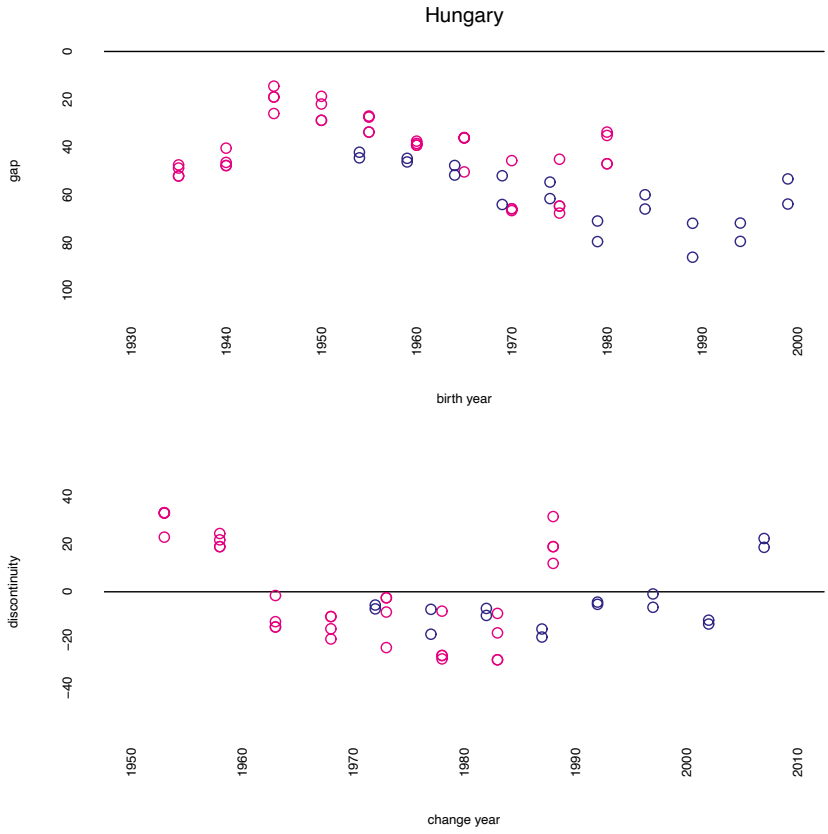


Figure 2.6e: Finland
(level by birth cohort above, 5-year discontinuity below)

Figure 2.6f: Hungary (level by birth cohort above, 5-year discontinuity below)

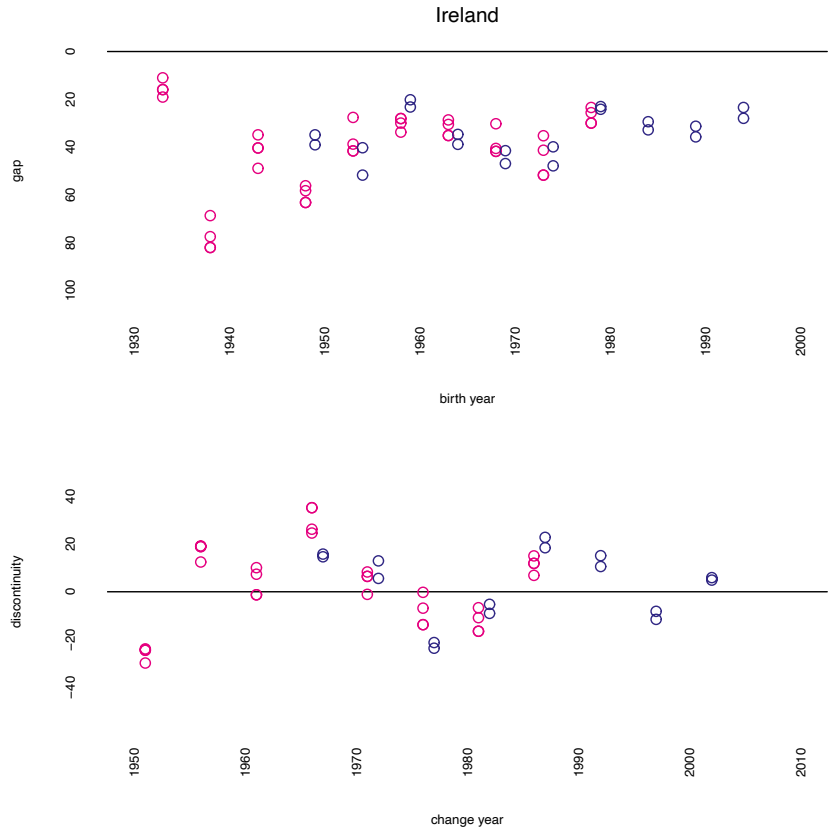


Hungary

The Hungarian results are not terribly promising for linking outcomes to policies. In both IALS and PIAAC, a change between the youngest age groups seems to be in the direction of improved equality of opportunity. Caution would suggest taking these results as age effects, as the apparent improvement oinf equality of opportunity in the late 1980s as observed in IALS is not confirmed by PIAAC, which in turn shows an improvement in the late 2000s. As noted above, this is the pattern that would arise if inequality of opportunity is lower when observed soon after the end of compulsory schooling than when observed a bit later.

The significant discontinuities in the 1950s seem more promising for linking improvements achieved to policies pursued. The discontinuities imply that cohorts born in the early and late 1940s did better in terms of equality of opportunity than those born in the early or late 1930s. This is possible, as the education legislation adopted in 1940 that very much fixed the later structure of Hungarian schooling was only implemented gradually from 1945 onwards.

Figure 2.6g: Ireland (level by birth cohort above, 5-year discontinuity below)



Ireland

In Ireland, IALS shows a pattern of improvement of equality of learning opportunity in the mid-1950s, followed, after a brief pause, by a more considerable improvement in mid-1960s.

Interestingly, Ireland also shows a very significant decline in equality of learning opportunity in the early 1950s. How this pattern should be interpreted is not clear, as the data set at our disposal is limited to a quite a small number of countries and only covers a few decades' worth of education policy, but if the pattern of initial improvement and later deterioration in equality of learning opportunity holds (see *Sequence of improvement and deterioration*), then this significant decline might very well be evidence of an earlier reform, some five to ten years before the decline in the early 1950s. And there is indeed such a reform, as the so-called 1943 reform of Ireland made primary education compulsory and affected cohorts born after 1931. Now the discontinuity estimated at 1951 in effect compares the 16-24-year-olds of 1961 to the 16-24-year-olds of 1951, and thus cohorts born before 1940 to cohorts born before 1935. Thus, the timing of the discontinuity would seem to match the 1943 reform quite well. If it is assumed that the Ireland of the 1940s and early 1950s is similar enough to the England, Northern Ireland and Ireland of the 1950s to show a very rapid succession of significant improvement in equality of learning opportunity followed by a significant decline, than the negative discontinuity in the very early 1950s would be matched by an improvement in the mid-1940s (1946+), which would correspond very nicely with the reform of 1943.

The timing of the first improvement in equality of learning opportunity in the mid-1960s corresponds to the reforms to comprehensive schooling in 1966 and the introduction of free secondary education in 1967. The improvement at the beginning of 1970 corresponds to the new curriculum of 1971 and extending compulsory schooling to age 15 in 1972, but the effects of reforms that follow each other in rapid succession cannot be separated with the very coarse data at our disposal.

The late 1980s and early 1990s also show an increase in equality of learning opportunity but this change may be too early to correspond to the Education Act of 1998, even if the PIAAC timing in effect compares the 16-year-olds of 2003 to those of 1993. If it is too early, we do not recognise any other education reform that would correspond to this increase in equality of learning opportunity.

Italy

For Italy, we may observe a significant improvement in equality of opportunity in the late 1950s and early 1960s. The timing of the improvement corresponds well to the reform of 1963, creating the unified *scuola media*, which lengthened compulsory schooling by three years, made lower secondary education free and abolished the previous system of tracking. In the preceding system the age group was divided between *ginnasio* and *scuola di avviamento professionale,* previous of which had entrance examinations and was academically and socially selective. This reform has been shown to be linked to both improved social mobility and increased educational attainment after compulsory education, and postponed the time of the first birth. (Flabbi, 1999; Fort, Schneeweis, & Winter-Ebmer, 2014) The timing of the effects of the 1963 reform has been estimated using numerous different data sources, and the cut-off cohorts separating the cohorts marked by improved enrolment in education and consequent educational attainment are usually born in 1950 or 1952, which matches our timing.

The improvement in equality of learning opportunity is followed by a gradual reduction of the improvement and then a decline in equality of learning opportunity ten years after the improvement, in the late 1970s. In Italy, this pattern of rise and decline is among the most drawn-out that we can observe in our data. This prolongation might conceivably be explained by the lengthy implementation of *scuola media*. It has been estimated that only in 1976 did the proportion of youth enrolled in lower secondary education approach 100 %. The lengthy implementation of the education reforms in Italy has been explained by lack of funding and the problem of dropouts due to the structure of the education system, which discouraged continuation in education when completion of the certificate was uncertain. The long aftermath might also be partially explained by the 1969 reform that widened access to tertiary education, allowing access to all secondary graduates and not only those with a qualification from *liceo*.

After the long aftermath of the *scuola media*, the next sharp change occurs in the late 1970s or early 1980s. By timing this corresponds to law 517 of 1977, which has been claimed to have 'established by far the most significant cultural and social changes in compulsory school, especially as regards teaching planning, monitoring and assessment of individual pupils, teamwork among teachers and co-ordination of the various classes, integration of disable pupils in ordinary classes and the consequent abolition of special classes.' (Garrouste, 2010)

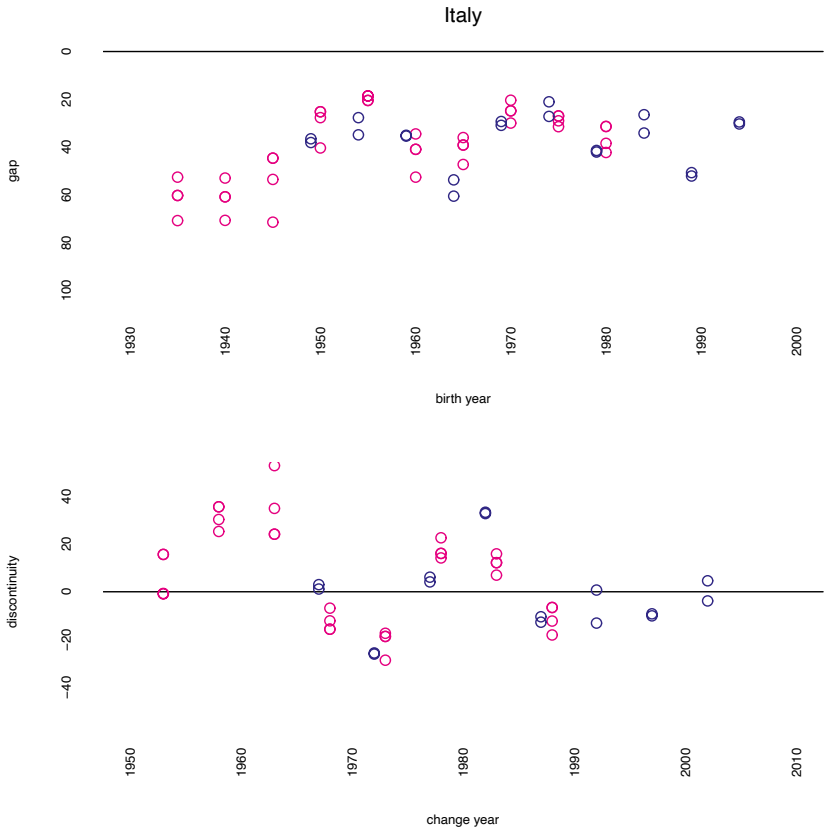


Figure 2.6h: Italy (level by birth cohort above, 5-year discontinuity below)

Netherlands

For the Netherlands, we can observe an increase in equality of learning opportunity in the early 1950s, which corresponds to the lengthening of compulsory education to 8 years in 1949, which was decided on in 1942, but only implemented in 1949. Here the Netherlands would match the experience in, e.g., Hungary.

Another, more considerable rise in equality of learning opportunity is apparent in the mid-1960s and thus falls very close to the timing of the Mammoth Act of 1968, which lengthened compulsory education to nine years and required children to attend six years of primary education and three years of post-primary education. It also unified the previously fragmented legislation of secondary education and has been estimated to have led to improved educational attainment even after compulsory education. (Plug, 2001). The improvements from the earlyish 1970s to 1980 are much smaller, but as IALS and PIAAC give a very consistent account of the improvements, we may note that they coincide with further changes in the 1970s, when the Compulsory Education Act of 1969 added a further year of compulsory education, requiring children to attend school between ages 6 and 16, for six years of primary and four years of post-primary education. This change in the length of compulsory education was implemented from 1973 and raised participation rates in 1975, as compared to 1974, and is often referred to as the Reform of 1975.

IALS shows a negative discontinuity in the early 1980s, but PIAAC does not show any sign of such a dramatic shift, which requires cautious interpretation. The decline in the early 1990s corresponds, by time, to the reform of 1992, which replaced the unitary junior secondary vocational education (LBO) with five different categories of schools.

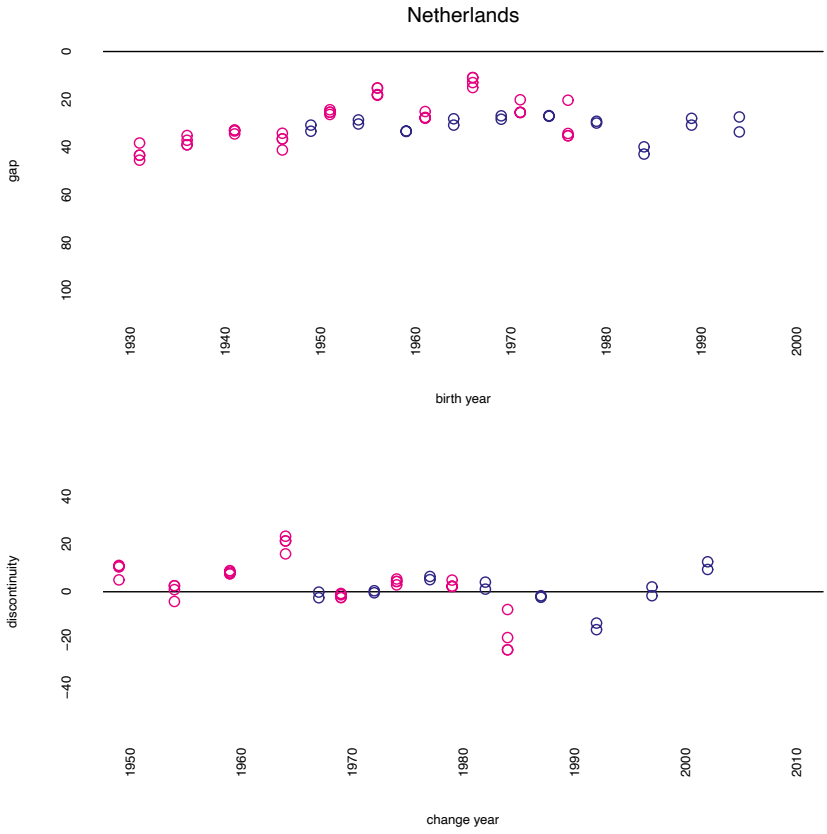


Figure 2.6i: Netherlands (level by birth cohort above, 5-year discontinuity below)

New Zealand

For New Zealand, the progression of equality over time is devoid of large discontinuities. As large portions of the last decades are characterised by a continuous decline in equality of opportunity, it is challenging to link that to policy changes, even on the basis of timing alone. The sole exception to the steady decline is the 1970s, which witness slight rises in equality of opportunity. The timing of the rise in equality of opportunity, which presents a break from the otherwise quite consistent decline over decades, coincides with the Private Schools Conditional Integration Act of 1975, which abolished many of the private schools as independent from the state by integrating them into the general school system and granting them public funding if they conformed to state standards and taught the state-approved curriculum.

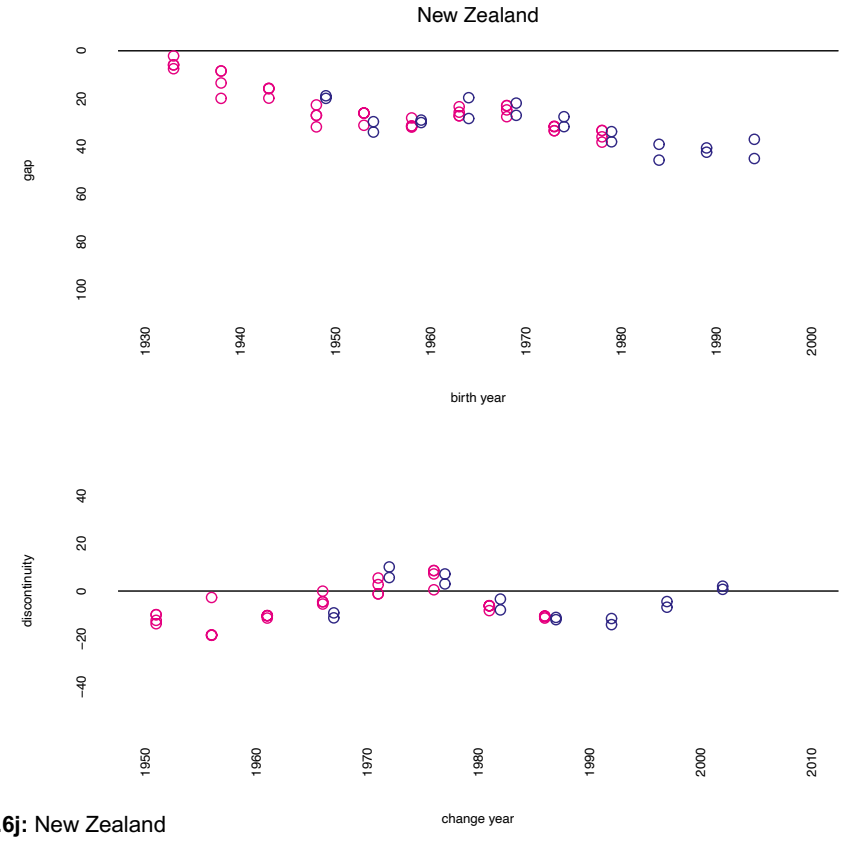


Figure 2.6j: New Zealand
(level by birth cohort above, 5-year discontinuity below)

Norway

For Norway, we can observe increasing equality of opportunity in the early 1950s, decreases in the early 1960s, relative stability from the late 1960s to the relatively late 1980s and a decline in the early 1990s. The initial rise in the 1950s is observed in many other jurisdictions as well, even though it is not universal, so it may feasibly be linked to general changes after the Second World War, rather than any specific reforms, as no policy reform takes place in the early 1950s.

The rises in equality of opportunity since the late 1960s seem to correspond – by timing – to the introduction of 9-year comprehensive schooling in 1969. Here the results would fit what we observe with similar reforms in other jurisdictions. The early 1990s show a drop in levels of equality of learning opportunity that would correspond to Reform 97, undertaken in 1997-1998.

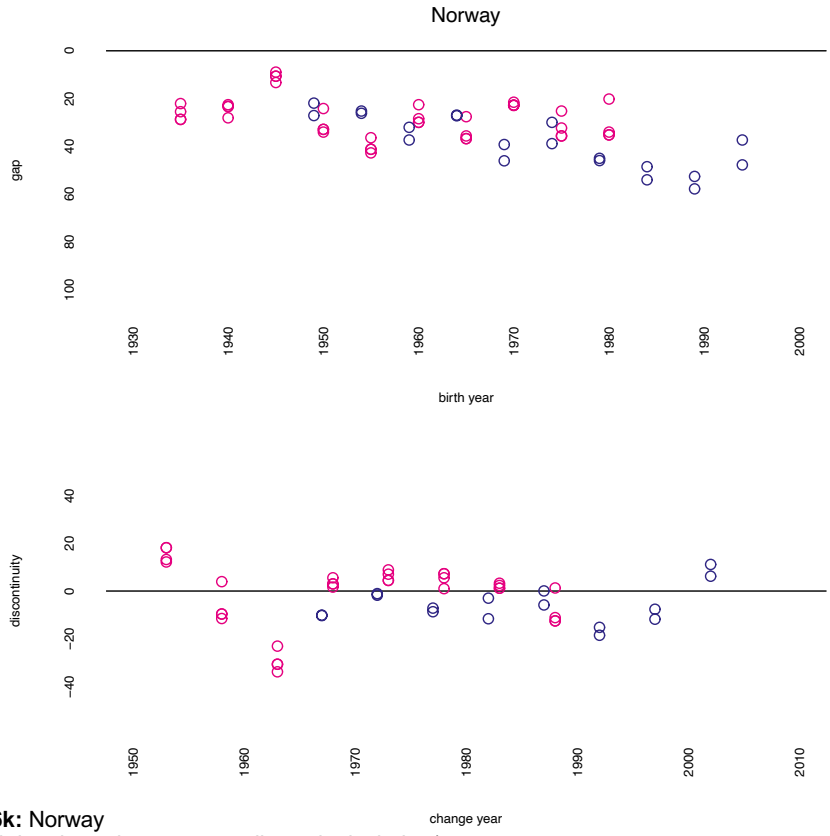


Figure 2.6k: Norway
(level by birth cohort above, 5-year discontinuity below)

Poland

In Poland, we witness rising equality of learning opportunity in the early and especially late 1950s, followed by stabilisation and decline. The very sharp increase in equality of learning opportunity at the very end of the 1950s would seem to correspond to the reform, implemented 1962-1966, that increased the length of compulsory education, among other things. This is one of the cases where it is useful to remember how the timing of the discontinuities is approximate and tries to match the timing of a potential reform by using only cohort data. As the analysis is performed with 5-year age bands, the positive discontinuity in the very late 1950s means that the potential reform that might have produced the reform did not affect the cohorts that became 16 years old in the very late 1950s.

Another, smaller increase is visible in the mid-1970s to mid-1980s, and a more notable decline in the early 1990s. The decline in the early 1990s is a common experience with many of the other post-communist countries, notably the Czech Republic and Hungary above. Given the link between income inequality and equality of learning outcomes, it seems possible that these changes are related more to general economic changes than to education policy. But here, the experience in Slovenia is different, as we shall currently see.

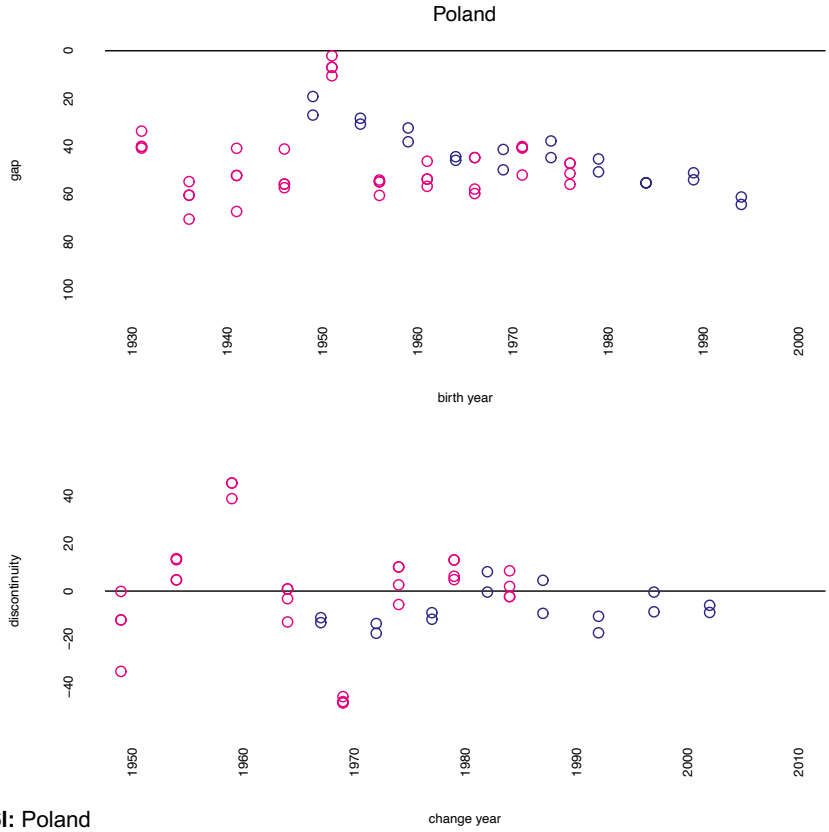


Figure 2.6l: Poland
(level by birth cohort above, 5-year discontinuity below)

Slovenia

Slovenia is one of the jurisdictions that experienced a rapid increase in equality of opportunity in the early 1950s. This development is quite close to the rapid increase in primary education enrolment in Slovenia right after the Second World War. The rise in the very early 1960s is, in turn, concurrent with the implementation of the extensive school reforms that were adopted in 1958 and came into force in Slovenia in 1959 and were implemented through the early 1960s. The period also witnessed increasing participation in secondary and tertiary education, with the government offering an increasing number of scholarships to individuals from poorer backgrounds. For the 1970s the results are rather stable, and there is an inconsistency between IALS and PIAAC. This inconsistency is also manifest in an apparent discontinuity towards higher equality of opportunity in the late 1980s that we can observe in IALS but not in PIAAC. Here, our preliminary interpretation would be that we are seeing an artefact produced by the age effect of the achievement gap rising somewhat between the youngest age groups from the late teens to early adulthood. This phenomenon is not unique to Slovenia, and it can perhaps be observed in PIAAC data as well, even though the overall effect is slightly smaller than in IALS.

Slovenia differs from the other post-communist countries (Poland, Hungary and the Czech Republic) in that we do not see a discontinuity in the direction of more inequality of opportunity in 1990s. This may be linked to Slovenia also experiencing very stable income inequality in the 1990s, in contrast to the other post-communist countries. The potential link between income inequality

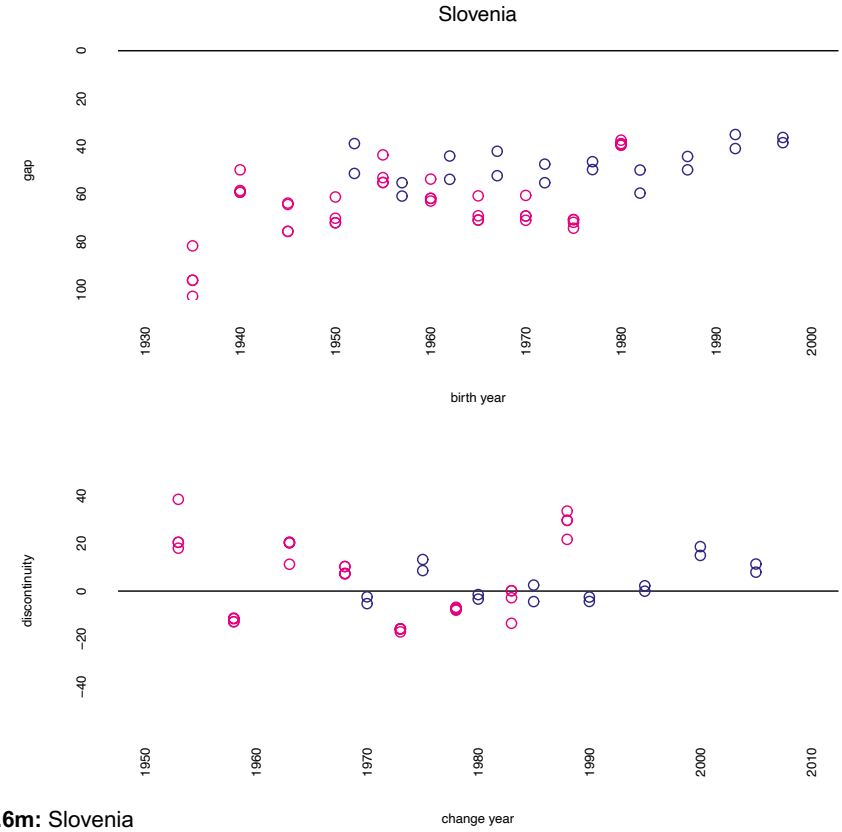


Figure 2.6m: Slovenia
(level by birth cohort above, 5-year discontinuity below)

would also seem plausible in the light of the previous analysis of the relationship between equality of opportunity and the level of learning outcomes, and the experience of Nordic countries, which mostly witnessed an increase in both inequality of opportunity and inequality of income distribution in the 1990s.

Sweden

For Sweden, the first significant rises in equality of learning opportunity occurred in the early and late 1950s. The comprehensive school reform is sometimes referred to as the 1950 Reform due to its first implementation beginning in the school year 1949-1950. The take-up of the reform progressed rather slowly across municipalities from the early 1950s but picked up towards the 1960s. This reform has been shown to have affected both educational attainment and earnings of groups with low parental education. For high parental education types, the results are more mixed, as their returns on education declined, but the overall population-level effect was positive. (Meghir & Palme, 2005)

A second significant discontinuity in equality of learning opportunity can be placed in the mid-1980s, but it is only visible in IALS and would seem to arise out of an age effect which also produces a positive discontinuity in the early 2000s in PIAAC. It should be noted that the discontinuities do not show traces of a significant change in the 1990s in Sweden, irrespective of the fact that in many respects – education policy reforms and growth of income inequality – the Swedish experience in and since the early 1990s has matched those of the other Nordics in our data, which display signs of a decline in equality of opportunity in early 1990s. Here our results thus agree with results that do not find increasing inequality after the Swedish school reforms of the 1990s. (Böhlmark & Lindahl, 2012)

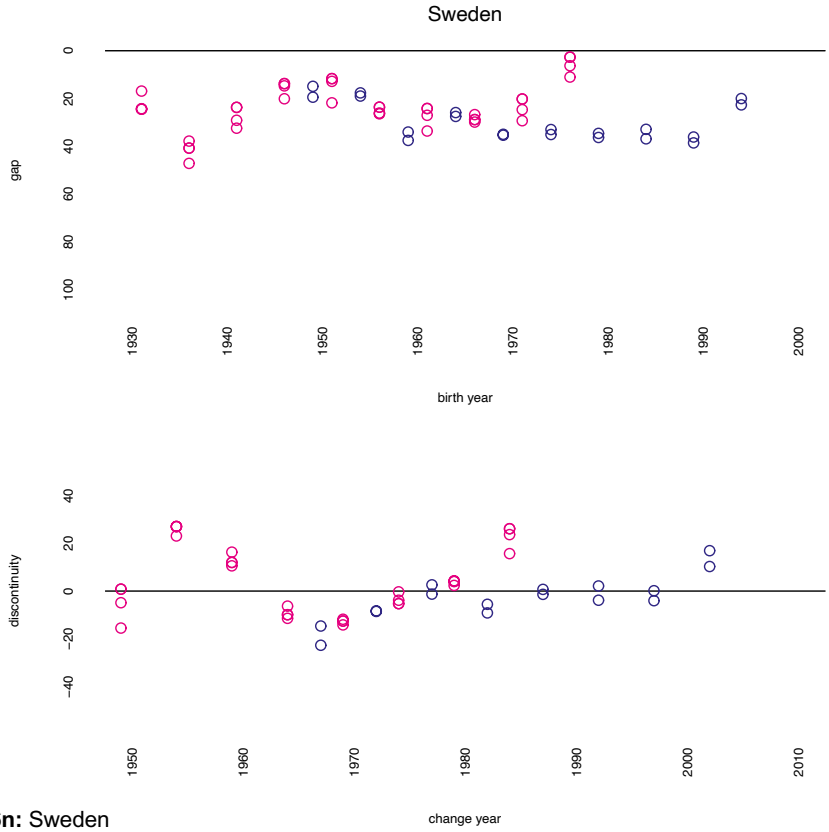


Figure 2.6n: Sweden
(level by birth cohort above, 5-year discontinuity below)

United Kingdom – England and Northern Ireland

For the United Kingdom, it is useful to look at England and Northern Ireland together, as there are important similarities in the policies pursued. These similarities may be relevant for the interpretation of the improvement in equality of opportunity in the late 1940s in Northern Ireland but not in England. England does not show an improvement, but rather a decline quite similar to the one observed in Northern Ireland five years later.

England expanded compulsory education in 1944, and Northern Ireland followed suit three years later. This might suggest that IALS captures the associated rise in Northern Ireland, with sufficient coverage of the pre-reform cohorts, but not in England, where the reform came into force earlier. Thus, the observed decline at the end of the 1940s in England would be equivalent to the decline in Northern Ireland in the early 1950s, and both could follow from a similar reform that initially improved equality of opportunity. This interpretation would fit the finding that the 1947 school-leaving age reform has been shown to have had a significant effect on cognitive skills, with an extra year of education improving specific measured skills by 0.5 SD. (Banks & Mazzonna, 2012) However, as IALS data does not provide good coverage of the pre-reform cohorts, this interpretation must remain tentative.

The sharp rise in equality of opportunity in the late 1950s does not seem to coincide with any reform, as the next significant reforms occur in the early 1970s, when compulsory schooling is ex-

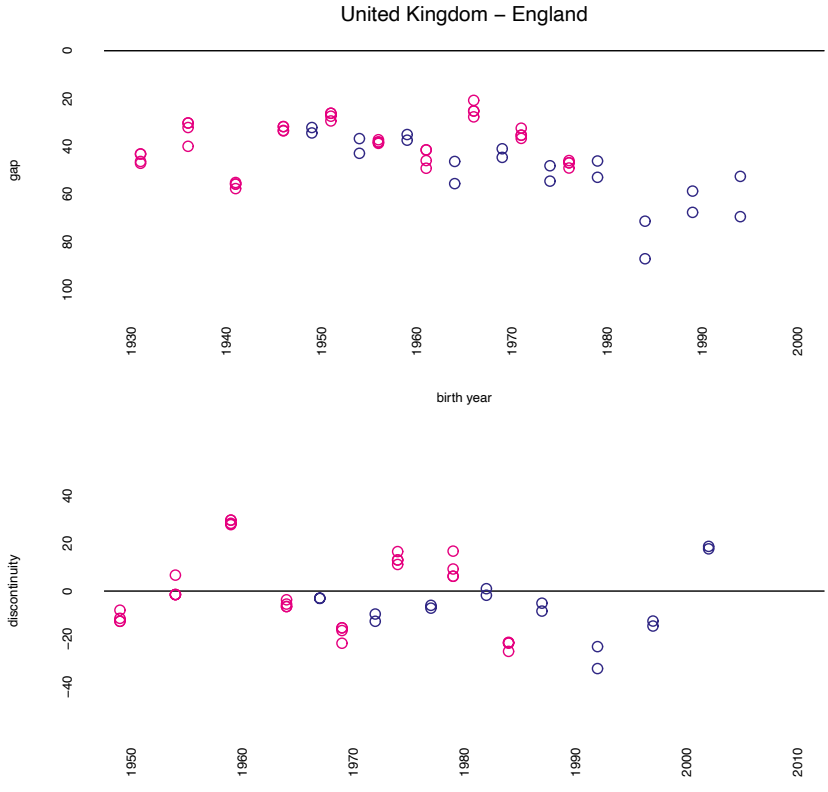


Figure 2.6o: United Kingdom - England
(level by birth cohort above, 5-year discontinuity below)

tended to age 16 in 1973 in England. However, this change coincides with rises in equality of opportunity only in IALS, while in PIAAC 1970 witnessed a decline in equality of opportunity in England. The change observed may also be linked to the fact that, through the 1970s, the inequality of learning opportunity in rates of staying in education after age sixteen increased (Blanden, Gregg and Machin 2005).

The early 1990s witnessed a sharp drop in equality of learning opportunity in both England and Northern Ireland, just as we have seen in many other jurisdictions. In England, the changes coincide with the School Standards and Framework Act of 1998.

The mid-to-end-1980s differ in England and in Northern Ireland, being associated with declining equality of opportunity in England and with increases in Northern Ireland. This difference in experience coincides with The Education Reform Order of 1989 in Northern Ireland and with the open enrolment reform of 1989 that widened access to grammar schools and improved educational attainment and equality of learning opportunity. (Maurin & McNally, 2007)

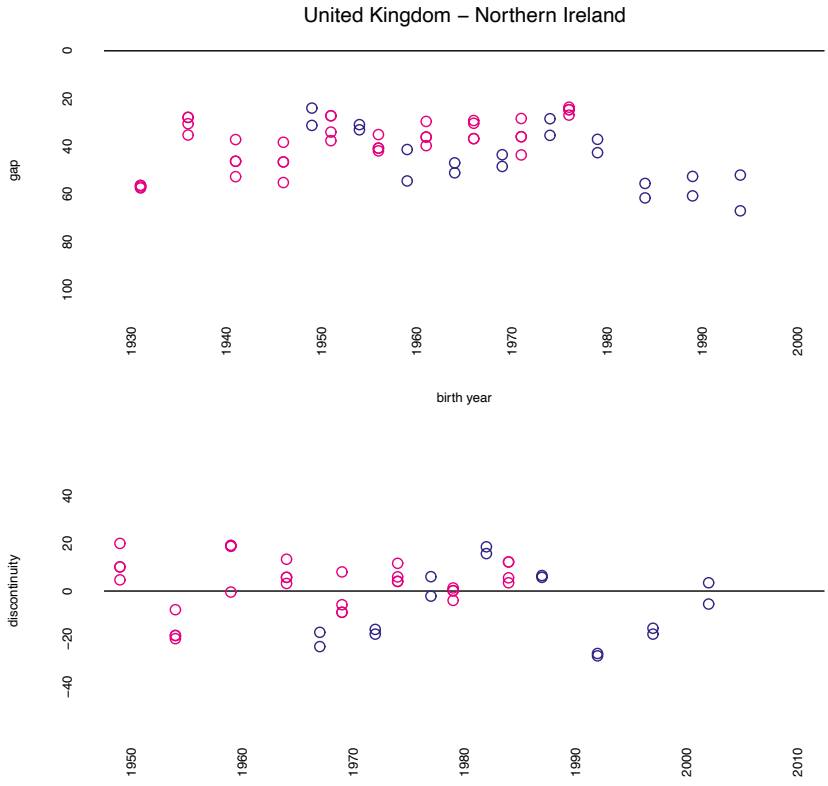


Figure 2.6p: United Kingdom – Northern Ireland
(level by birth cohort above, 5-year discontinuity below)

12.3 Reforms by type and a closer look at Finland

As noted above, IALS and PIAAC give a quite consistent picture of the level of equality of learning opportunity for most jurisdictions most of the time. Discontinuities are more difficult, as they require more granularity and the risk of overinterpretation increases when we approach the limit of the resolution of the data. Fortunately, the 5-year discontinuities from both surveys match reasonably well for many jurisdictions, which gives some confidence that we are not simply imagining things even if we may be seeking weak signals.

Reforms by type

Many of the jurisdictions present in the sample have made similar reforms over the decades covered, and the timing of many of the reforms corresponds to discontinuities in equality of opportunity. Obviously, education policy is not conducted in isolation, and reforms may coincide with substantial social changes. That is why we have highlighted changes related to, e.g., the end of communism in Eastern Europe. But where education policy reforms have an effect, similar reforms could, in general, be expected to be associated with similar results. Thus, one way to interrogate the results obtained with roughly matching discontinuities and reforms by timing is to see if similar reforms are associated with similar discontinuities – or at least those in the same direction.

Our analysis is mostly limited to a selected set of developed countries that have been relatively developed for most of the time period of analysis, and thus at least the lower extremes of variation of contexts have been removed from the analyses. Despite this, it should be clear that the contexts of the Nordic countries (Denmark, Norway, Sweden, Finland) are not interchangeable with Anglo-Saxons (England, Northern Ireland, Ireland), Southern Europe (Italy) or Eastern Europe (Czech Republic, Hungary, Poland).

A few reform types or events emerge as covering most of the identified reforms that are linked to discontinuities, mostly improvements, in equality of opportunity. The 29 discontinuities linked to education reforms or social changes by their timing are presented in ***Table 3***.

A significant number of discontinuities in levels of equality of learning opportunity are associated with changes in the length of compulsory education. This is consistent with earlier research, which shows extensions of compulsory schooling being associated with decreased socioeconomic differences. Extension of compulsory schooling would seem to be associated with improved equality of learning opportunity in Denmark, Finland, Hungary, Ireland, Italy, Netherlands, Norway, Poland, Slovenia, England and Northern Ireland. Reduction of compulsory schooling is associated with a decline in equality of opportunity in the Czech Republic.

Another type of reform that is often associated with discontinuities in equality of learning opportunity levels are comprehensive school reforms in the 1950s-1970s. These reforms seem, in line with earlier research, to be associated with increased levels of equality of learning opportunity. Nordic comprehensive school reforms are associated with improvements in equality of opportunity in Denmark, Finland, Sweden and Norway. Similar reforms in England, Northern Ireland and Ireland seem to be linked to improvements as well, as are the reductions in tracking with unified secondary schooling in Flanders and Italy. The move towards a more unified system with private school integration in New Zealand would also sit broadly in this category of unifying the education system. Correspondingly, increased fragmentation of secondary schooling in the Netherlands is associated with decreased equality of opportunity.

Table 3: Reforms and discontinuities

Jurisdiction	Reform	Time of reform	IALS	PIAAC
Belgium - Flanders	Schulpakt / Pacte scolaire	1958	X	
	Unique secondary school	1971	X	-
Chile	Eduardo Frei Montalva	late 1960s	X	X
	Democratisation	1990	X	X
Czech Republic	Compulsory schooling reduced	1979	X	X
	End of communism	early 1990s		X
Denmark	Compulsory schooling extended to 9 years	1971	X	X
	School reforms	1993-1994	X	X
Finland	Compulsory schooling extended	1957	X	
	Comprehensive school	1970s	X	X
	School reforms	1994-1998		X
Hungary	Education legislation of 1940	Turn of 1940s & 1950s	X	
	End of communism	early 1990s		X
Ireland	Compulsory primary certificate	1943	X	
	Comprehensive school, free secondary education, new curriculum and extension of compulsory education.	1966-1972	X	X
Italy	*Scuola media*	1963	X	x
	Law 517	1977	X	X
Netherlands	Compulsory schooling extended to 8 years	1949	X	
	Mammoth Act	1968	X	X
	Abolition of LBO	1992		X
New Zealand	Private Schools Conditional Integration Act	1975	X	X
Norway	Comprehensive school	1969	X	X
	Reform 97	1997		X
Poland	School reforms, extension of compulsory schooling	1962-1966	X	
	End of communism	Early 1990s		X
Slovenia	Enrolment increase + reforms of 1959	early 1960s	X	
Sweden	Comprehensive school	1949-1962	X	
United Kingdom - England	Compulsory schooling extended	1944	X	
	Compulsory schooling extended	1973	X	
United Kingdom - Northern Ireland	Compulsory schooling extended	1947	X	
	Education Reform Order	1989	X	

In contrast to these changes, most often associated with initial increases in equality of learning opportunity, some countries seem to display discontinuities in declining equality of learning opportunity in the 1990s. Here we have two types of jurisdictions where such discontinuities are observed. On the one hand, we have the Nordics – Denmark, Finland and Norway – and on the other the post-communist countries. One potential commonality between post-communist and Nordic countries is an increase in income inequality, especially as Slovenia acts as a counter-example that shows no signs of increase in either inequality of opportunity or of income. Thus, there is a challenge in linking these changes to changes in education policy, rather than to more general changes in society. The Nordics implemented very similar reforms, emphasising choice, differentiation and local autonomy, but because of the other social developments the case for the causal role of these reforms is far from clear. For one, Sweden does not experience a similar increase in inequality as the other three, when we use adult data. Nonetheless, it may be noted that the Nordics implemented reforms that emphasised decentralisation, and we saw above that changes in the other direction – towards less tracking and differentiation – often lead to increasing equality of opportunity. Thus, reforms leading to differentiation may perhaps lead to opposite results from the changes towards more comprehensive systems.

A closer look at Finland

The reason for taking a closer look at Finland is the fact that our analysis shows evidence of the types of reforms implemented in many other countries. Finland has made changes in the length of compulsory education as well as reforms towards a more comprehensive schooling system and later changes towards a more individualised or differentiated system. The results suggest that the extension of compulsory schooling in the late 1950s and the introduction of comprehensive school in the 1970s were associated with increased equality of opportunity. More interestingly, Finland may be of particular interest as it has acquired a worldwide reputation as a leading educational system. Some of the popular interpretations of the reasons behind Finnish success have raised the reforms of the early and late 1990s to the fore. Our results, in contrast, would seem to suggest that the reforms of the 1990s are associated with an exceptionally sharp decline in equality of learning opportunity.

As with other jurisdictions, IALS and PIAAC give a very consistent picture of the development of equality of learning opportunity in Finland across time. The evolution of the achievement gap in the light of IALS and PIAAC also matches the results from youth large-scale surveys quite closely, as we can see by comparing Finland in ***Figures 2.5a*** and ***2.5b*** to the high-low parental education achievement gap reported in Figure 9.2 of Salmela-Aro and Chmielewski (2019). Both sources of data provide evidence of a curvilinear evolution of equality of opportunity over time in Finland, with first a decreasing and then increasing achievement gap. This curvilinear evolution of the achievement gap also makes Finland one of the countries where Chmielewski (2019) does not observe a rising or declining trend from the earliest cohorts born in the 1950s to the youngest born in the early 2000s. There is a slight difference in timing between the two time-series, though. While the youth data imply an uptick in the achievement gap in cohorts born in the early 1990s, adult data place it somewhat earlier, in the mid-1980s. This result is perhaps not surprising, as differential participation in general upper secondary and in tertiary education might be expected to increase the achievement gap of measured skills from what was observed towards the end of compulsory education. Thus, if policy changes on several levels of the education system, we might expect to see evidence of the reforms in older cohorts in adults who have been affected only by changes in tertiary or upper secondary, rather than when we observe the outcomes in youth, as in many large-scale surveys.

We have noted above that an improvement in the late 1950s coincides with an extension of compulsory education and another in early 1970s with the implementation of the comprehensive school reform. The results suggest that the introduction of nine-year comprehensive school in 1972-1977 led to relative gains in the learning outcomes of lower socioeconomic groups. These relative gains were shown by Uusitalo and Pekkarinen (2012) to persist into early adulthood, and our results suggest that they persist into mature adulthood as well, translating into smaller achievement gaps.

In the early 1990s and early 2000s, where PIAAC shows a significant decline in equality of learning opportunity, the start of the decline would correspond to changes in the core curriculum in 1994 and the legislative reform in 1999, which increased individual/parental choice in orientation of studies with increased streaming in schools, school specialisation and school choice. The high level of institutional autonomy and teacher autonomy was introduced into the system in the reform of the Core Curriculum of 1994 and the reform of education legislation in 1999. Simola (2015, 2017) has argued that Finnish education policy before the reforms of the 1990s was characterised by distrust of teacher autonomy and a view that teachers formed an obstacle to the implementation of policy. Trust was introduced to the system during and after the deep economic recession of the 1990s. (Aho, Pitkänen, & Sahlberg, 2006; Sahlberg, 2009; Simola, 2017) The timeline of the evolution of equality of opportunity would suggest, especially when we note the association between equality of opportunity and learning outcomes, that the role of the reforms of the 1990s in the Finnish success is less clear than many may have believed.

The reforms of 1999 introduced increased individual/parental choice in the orientation of studies with increased streaming in schools, school specialisation and school choice. In terms of governance, this was implemented through reduced administrative regulation and oversight, a move to lump-sum funding and an overall increase in local autonomy, underpinned by an emphasis on the local understanding of local conditions and on local democratic accountability. These reforms have been linked to growing socioeconomic differentiation in Finnish schools, observed as early as the turn of the 1990s and 2000s. (Bernelius, 2008, 2013; Kosunen, 2016; Seppänen, 2006; Seppänen, Kalalahti, Rinne, & Simola, 2015)

The link between growing socioeconomic differentiation and decline in the level of equality of learning opportunity links the results to earlier research that shows early institutional differentiation negatively affecting equality of learning opportunity by increasing the information requirements to navigate successful educational careers. (Gamoran & Mare, 1989; Pfeffer, 2008, 2015; Woessmann, 2009) To some degree, this differentiation has gone unnoticed as it has happened within schools. In comparison with other Nordic countries, notably Sweden, Finland has seen between-school differences remaining rather small but large class effects rising within schools. (Hansen et al., 2014)

IALS and PIAAC also give a similar timeline in terms of the development of proficiency in Finland. In IALS and PIAAC, the 5-year cohorts with the highest proficiency were born before 1978 and 1982, having thus finished their compulsory schooling by 1994 and 1998, respectively. Longitudinal data from cognitive tests show an anti-Flynn effect of declining IQ in Finland beginning in 1997 for the age group 18-20, corresponding to the cohort born 1977-1979 that finished compulsory schooling in 1993-1995 (Economic Policy Council, 2018; Koivunen, 2007)

Also, recent work conducted in the Finnish Education Evaluation Centre (FINEEC) has made national assessments of the learning outcomes of mathematics comparable over time. The results show that the highest level of achievement was recorded in 2001, after which the results have undergone a decline of very much the same magnitude as has been observed in PISA between 2006-2015. These results in the direction of development of proficiency are in accordance with the above-reported results that equality of learning opportunity is associated with the level of proficiency. Overall, the

cross-sectional data provided by IALS and PIAAC on Finland is consistent with the picture painted by longitudinal data in the levels of both equality of learning opportunity and proficiency, and several different sources of data suggest a turn towards greater inequality of learning opportunity taking place in the mid-1990s.

As noted above, this turn coincided with reforms that have been associated, in popular accounts, with Finnish success in international surveys, notably PISA. The timeline that has been emerging since the first rounds of PISA suggests that some other changes, often credited for Finnish success, may also need careful empirical analysis. For example, the master-level education of Finnish teachers started in 1979, with the first cohorts of the new teacher education graduating in the mid-1980s. This implies that the cohorts with the highest level of equality of learning opportunity and proficiency began their compulsory schooling when the new master-level teachers were just beginning to be introduced into the teacher force, and finished their compulsory schooling while no more than a third of all active teachers had master's degrees. Thus, if master-level teacher education has a direct effect on equality of opportunity or the level of learning outcomes, its significance must either be drowned by other causes, or it must be very nonlinear.

Our results provide some indications on possible directions for future research. In the case of Finland, the reforms in the 1990s seem to be associated with a decline in the level of equality of learning opportunity that is even more significant than the increase in equality of learning opportunity associated with comprehensive school reform of the 1970s. Our results are consistent with the interpretation of Sahlberg (2007, 2011, 2015) that in the history of Finnish education, high levels of proficiency are associated with high levels of equality of learning opportunity. However, they would imply that the causes, namely teachers, trust and autonomy, credited by, e.g. Sahlberg and Simola (2015, 2017) for the level of outcomes of the Finnish school, might merit reconsideration.

The association of Finnish education policy reforms of the 1990s with decreased equality of learning opportunity and proficiency have several implications for the interpretation of the trajectory of the Finnish education system. They would imply that the recently observed decline of Finland in PISA is an aspect of a longer-term trend of declining proficiency and equality of learning opportunity. It also raises the possibility that the apparent concurrent decline of proficiency and levels of equality of learning opportunity undergone by many (mainly European) countries might be not only interlinked but also linked to changes in the direction of the education policy in the 1990s, rather than to forces beyond the influence of policy. However, much more research is needed, as there are several additional factors at play. For example, while it may be noted that many Nordics see increases in inequality of opportunity in and since the early 1990s, they have not only experienced changes in education policy, but also significant changes in income inequality in the 1990s.

Caution

The method used above does not allow us to assert that the associations observed are causal, even if similar reforms tend to be associated with similar equality of learning opportunity discontinuities in different countries. It is also possible that similar education reforms have been systematically associated with other policy reforms that are causally responsible for the discontinuities, or the association is explained by similar reforms having been implemented in similar times. Further research is needed to settle the issue.

The method for observing age group discontinuities consistent with specific education reforms also has significant limitations due to the roughness of the data. Due to low-resolution data, the association between reforms and discontinuities cannot be very exact and, in cases of frequent reforms, the resolution of the data is quite insufficient. For example, the Czech Republic changed its length of compulsory education three times in little over a decade, with lengthening (1948, 1960) from eight to nine years and shortening (1953) from nine to eight years in quick succession and Italy implemented several education reforms in the 1990s and 2000s, with large discontinuities in equality of learning opportunity at the level of 5-year age groups, but no clear trend at the level of 10-year age groups. The lack of resolution also makes it difficult to interpret the results for cases such as Germany and the USA, where states have far-reaching powers in education, in effect separating a jurisdiction into several education systems.

Research is also needed to form a consistent picture from adult data and student assessment data. For example, both widely-used time series of education quality based on large scale assessment of youth (Altinok, Diebolt, & Demeulemeester, 2014; Hanushek & Woessmann, 2015) date the start of the decline of outcomes in Finland to the early 2000s, while adult data would suggest an earlier beginning for the decline. A reasonable hypothesis might be that any policies or other changes that affect inequality might also affect through upper secondary and tertiary education and would thus have, through them, effects on the cohorts that went through compulsory schooling unaffected by the changes. But that is a hypothesis that would need to be verified.

Summary

The results suggest that observed differences between age groups in adult data can be, with caution, used to assess differences between birth cohorts. For the set of 18 jurisdictions in our sample, level of equality of opportunity in different cohorts was quite consistent in IALS in 1994-1998 and in PIAAC in 2012-2017, despite the 14-21-year gap between the observations. This suggests that even if the level and perhaps the distribution of skills may evolve in adulthood, the basic pattern of distribution of observed ability is stable enough to preserve differences in the relative distributions in different cohorts.

There is more uncertainty in terms of the discontinuities between cohorts, but IALS and PIAAC also provide a largely consistent picture for the discontinuities. Moreover, many significant discontinuities seem to be associated, in time, with policy changes. The fact that similar reforms are apparently often associated with changes in the same direction – one that has sometimes been confirmed by previous research to be associated with the reform in question – suggests that what is observed may be more than overinterpreted noise.

SECTION VI - SUMMARY AND DISCUSSION

13 Drawing strings together

In this study, we have approached the 'fundamental question in sociological research on education and a primary concern of educational policy-making' of 'whether socio-economic equality in educational opportunities can be increased without lowering the quality of education' (Pfeffer, 2015) through the keyhole of adult skills.

The limitation of available data and the observational nature of the data that is available have highlighted the need not only to explicate the limitations imposed on the analyses by the limitations of the data, but also to find conceptual solutions to do justice to the objectives of the research while recognising the limits of what can be empirically achieved.

13.1 Theoretical contributions

Socio-economic equality, opportunity and the equity-efficiency trade-off

While the potential trade-off between socio-economic equality and quality of education has been recognised by Pfeffer, among others, the potential origins of such a trade-off have been less well explicitly recognised.

It has been recognised that equality of opportunity does not form a trade-off with efficiency, but rather contributes positively to efficiency. Inequality of opportunity entails losses in efficiency or excellence, as resources needed to realise potential are allocated in a way that leaves some actualisation of potential short of what would be possible and places those resources where their marginal product is smaller. Thus, inequality of opportunity would not, in theory, be expected to be in a trade-off with excellence or the level of educational outcomes.

Recognition of the difference between outcome and opportunity trade-offs allows us to recognise that the potential trade-off between 'socio-economic equality in educational opportunity' and the quality of outcomes does not arise if opportunity is indeed opportunity – access to resources needed to realise potential – but only if observed circumstances of socioeconomic background are associated with innate ability.

The standard approaches to equality of opportunity infer inequalities of opportunity from differences in expected outcomes between types defined by their circumstances. Variation of outcomes with circumstances beyond individual control is taken to be unjust inequality of opportunity. If innate ability is not taken to be a circumstance and an unjust reason for learning outcomes, then the standard approach to identifying inequalities of opportunity is vulnerable to unobserved selection of type by innate ability.

Thus, the Pfefferian fundamental question only arises when social hierarchies are matched by hierarchies of natural ability so that the achievement gaps between socioeconomic types are explained by unobserved differences in innate ability, rather than the observed differences in circumstances. If that is the case, limiting the socioeconomic achievement gap does not entail equalisation of opportunity, but equalisation of outcomes, which we know to form a trade-off with education.

The theoretical section above contributes to clarifying that the difference between outcome and opportunity trade-offs needs to be recognised, to better recognise that the Pfefferian fundamental

question is nothing less than a question on whether observed social hierarchies are a direct reflection of an underlying natural hierarchy of ability or not.

Innate ability and merit

The question of socioeconomic differences in innate ability is closely associated with several theoretical and empirical questions that are arguably of increasing interest as cognitive achievement becomes increasingly important in the establishment of social hierarchies. While we have limited the scope of this work to intellectual cognitive achievement[24], the social and emotional aspects of cognition are arguably not fundamentally different in this respect. After all, personality traits seem to be partially biologically determined, and thus the questions related to rewarding individuals for achievements beyond their control become relevant. For personality traits and socio-emotional skills, the same distinction between more immediate phenotypic outcomes, such as learning outcomes or displayed personality, and more ultimate welfare outcomes, may be even more relevant than for intellectual cognitive outcomes. Even if personality were fixed entirely by genetics, the rewards for different traits would be a question open to social consideration and normative discussion.

The challenge is that, while innate differences in individual potential are widely recognised, the concepts of responsibility and merit are intimately linked to freedom and the need to bear responsibility for the consequences of one's exercise of freedom of choice and action. The unavoidability of recognising individual differences in aptitudes, and the moral arbitrariness of those aptitudes, makes it necessary to reject either attempts to morally justify merits or to hold individuals morally responsible for causes beyond their control. This has introduced considerable tension into the philosophical positions of many traditions. While entitlement theory has readily accepted innate ability as a just cause for entitlement, it has not fully faced up to the implications of this stance for the defence of individual responsibility as a consequence of liberty. If individuals are to be held responsible for causes beyond their control, then responsibility cannot be grounded on freedom. It is entirely unclear what responsibility would be based on, if it were be based on robust causal responsibility.

Alternatively, the need for a moral justification of rewards can be rejected altogether, but then we step beyond philosophy into the area of assertion or divine revelation. A refusal to justify rewards on moral terms, or an assertion that they do not require moral justification, amounts to demanding the acceptance of principles that have not been justified by argument providing a basis for them. This may do for practice, but it will hardly do for theory or philosophy, which require assertions to be based on something other than revelation. Dworkin (2011) notes that much moral philosophy is born from the critique and defence of proposed principles of justice. Moral philosophers:

> do not say: 'It is too bad that our principle has such consequences but that is how the cookie crumbles. Our principle is just, as it happens, true.' We would be appalled if they did: it makes sense to ask for support even for a very abstract moral principle, and in some circumstances it would be irresponsible not to try to provide one.

For the Rawlsian approach, the tension introduced by the importance of innate characteristics to outcomes is apparent in the way Rawls would want to treat innate ability as a natural resource, the benefits of which should be distributed to all equally, while at the same time holding innate ability to be, via Fair Equality of Opportunity, a legitimate basis for selection into offices and positions to which

[24] Socio-emotional traits or skills are not non-cognitive, even if they are sometimes popularly labelled as such.

social and economic inequalities may be attached. The former approach rejects rewarding individuals for their innate characteristics, the latter approves of – and provides a mechanism for – it.

We have proposed a conceptual solution to the challenge, introducing the concept of a natural responsibility characteristic. It refers to causes of outcomes that are considered morally just by virtue of being unavoidable, thus justifying individual moral responsibility in the absence of causal responsibility. This allows us to consistently hold individuals responsible for their innate aptitudes when it comes to outcomes unavoidably linked to those aptitudes, be they physical height, personality or cognitive achievement. At the same time, it allows us to reject innate characteristics as a sufficient justification of economic, status or welfare advantages, as the link between innate abilities and such more down-stream outcomes is more socially malleable.

This approach relies on the distinction between fundamental principles of justice and rules of regulation as an implementable version of those fundamental principles in a given situation. It recognises that, while it may be fundamentally unjust to hold an individual responsible for something beyond their control, sometimes we do not have a choice. When we have or gain such a choice, we have to justify the outcomes using moral reasons. This links morals, as rules of regulation, not only to the specifics of our existence in terms of natural laws of this universe and our nature as naked bipedal primates with opposable thumbs, but also to the level of technology available to us. After all, anything that follows for the rules of regulation from our characteristics being different from those of, e.g., a consciousness enveloping a whole planet (Lem, 2012), follows from the technology available to us. With sufficient technology, the hypothetical intervention bridging that difference is available to us, while without it the intervention is neither well-defined nor implementable.

Opportunity and observation

On a related note, we propose embracing ill-defined causation for the purposes of social science when we try to approach a causal phenomenon that is so complex that full mechanistic causal understanding is beyond our capacity. This proposal arises from the need to acknowledge equality of opportunity as a fundamentally causal concept, and that the empirical approaches to measuring it amount to causal inference with the causal relations being provided by a presupposed causal model, while the nature of the exercise is often obfuscated by describing it only in technical terms as measuring inequality through the covariation of outcomes with observed circumstances.

It has been emphasised in the literature that measuring equality of opportunity needs a correctly specified causal model, and the literature recognises that the full mechanistic understanding of such processes as intergenerational mobility or equality of opportunity is beyond our abilities. Our contribution is to provide language that combines the recognition of both facts with the view that we should be explicit in recognising the causal aims of the research when they may be difficult to reach.

Embracing the concept of ill-defined causality, together with the use of causal diagrams, would provide clarity to the nature of the research into complex social phenomena. For example, it would facilitate recognition of the fact that many of the measures used in social science are not to be understood as well-defined treatments and treating them as such may easily lead us astray. For example, indicators of the socioeconomic position should at this stage not be understood as well-defined treatments, so that the implicit treatment in the difference between individuals with high and low parental education is the explicit treatment of providing parents a formal qualification. Instead, they should be more explicitly understood as ill-defined treatments that we do not yet understand on the level of

well-defined treatment. The implicit treatment may be partially genetic or biological, partly social and economic.

Recognising the difference between ill- and well-defined treatments would be helpful in avoiding treating ill-defined treatments as if they were, or were close to being, well-defined ones. Accepting that indicators of SES are ill-defined treatments would allow us to explicitly accept that the treatment may have a biological or genetic component and is not necessarily entirely environmental even though it is defined by an environmental proxy, such as parental education or income.

Now, the risks involved with misdiagnosis of causes of distributions are often not symmetrical. While misdiagnosing innate and non-malleable causes as environmental risks inefficiency, as resources are used to treat an untreatable condition, misdiagnosing environmental causes as innate risks condemning individuals – without choice or fault of their own – to inferior outcomes as no effort is made to treat a treatable condition. Thus, in the field of policy, hereditarian and environmental positions are not equally relevant when there is uncertainty as to the relative merits of the two positions. Here the principles of distributive justice meet, as they often do, the principles of retributive justice. Assumptions of innocence until proven guilty or guilt until proven innocent are not symmetrical and should not be taken as such.

Natural and social inequality

Ill-defined causality may help when social science needs to come to grips with the implications of behavioural genetics. At the moment, our understanding of the causal link between genotype and behavioural phenotype is very much incomplete. Whether our understanding will ever become complete enough for us to know the full phenotypic opportunity set of an individual genotype in a range of environments is still an open question. For now, we surely do not possess such understanding. This is significant in social science, as it is possible that social differences are grounded in biological differences that make them unavoidable. The challenge is that, while there has historically been no shortage of assertions that social inequalities simply reflect underlying natural inequalities and thus those in a worse position have earned their lot, the evidence for such assertions is far from conclusive. Unfortunately, heritability estimates are still sometimes – even quite often – used as measures of the biological fixity of a trait, a task to which they are not suited.

These age-old questions become relevant as behavioural genetics is increasingly able to find biological genetic correlates for behavioural differences. We know – and have known from the twin study era – that all behaviour is heritable. The age of molecular behavioural genetics is increasingly revealing this with a direct association between genetics and behavioural phenotypes. However, we also know that heritability is not a measure of the malleability or unavoidability of the phenotypic outcome. The genotype is in constant interaction with the environment, and highly heritable traits have shown great responsiveness to environmental changes.

The fundamental moral, empirical or political questions of social science are causal and counterfactual in nature. They ask why things are the way they are. To what degree, and under what conditions, could they be different? Heritability does not answer these questions. Indeed, it does not even answer the fundamental question of behavioural genetics, the history of which can be understood as an attempt to progress from weak biological explanation to strong biological explanation. This has not prevented some from declaring group differences as unavoidably grounded on biological differences, based on heritability estimates. Now that behavioural genetic data is increasingly available,

and it is widely recognised that humans are not blank slates, it would be essential to prevent old fallacies from creeping back in through the back door.

Acceptance of ill-defined causality would come in useful here, as it would allow a more explicit description of the causal mechanisms postulated by different models. It is useful to recognise that the opportunity trade-off only arises if socioeconomic groups differ by innate ability. Indeed, there is a long history of justifying sex, race and social inequalities through the non-malleable differences between sexes, races and social classes in intelligence. This history should not prevent researchers from asking and exploring such questions. It is possible that such differences have a biological, or even unavoidable, basis, even if assertions that they do may traditionally have been based more on prejudice than on evidence. But such hypotheses should also be explicitly recognised, as it allows them to be better assessed against empirical findings. For example, it can be hypothesised that increasing socioeconomic differences in cognitive achievement and a decline in the achievement of lower socioeconomic groups occurs with increasing heritability of cognitive outcomes. This may be the case, but it entails a hypothesis that, in the status quo ante, the level of socioeconomic differences in achievement have been narrower than what they would naturally be. The mechanisms for this could be, e.g., additional instruction for cognitively or economically poorer pupils. In general terms, this hypothesis implies that, while socioeconomic groups differ quite significantly in their resources, the distribution of social resources is narrower than that of innate ability. This is, by historical standards, a rather extreme position in terms of social differences. While many have defended the view that social hierarchies replicate natural hierarchies and should thus not be interfered with, it has not been generally asserted that social inequalities are actually smaller than natural inequalities.

13.2 Empirical contributions

The empirical sections on equality of opportunity and learning outcomes relate to the causal understanding of equality of opportunity on the one hand, and to a better understanding of the phenomenon to be explained on the other.

The first half of the empirical analyses focuses on the fundamental question of a trade-off between socioeconomic inequality, the nature of the association of equality of opportunity, and the level of learning outcomes. This part does not seek to establish causality, but rather to clarify the nature of a phenomenon to be causally explained. It seeks to understand to what degree socioeconomic differences are associated and co-evolve with the level of learning outcomes. Establishing a link between the two does not entail establishing a causal relationship between the two, with one causing the other, but rather establishing something about the shape of the phenomenon to be causally explained.

The second section of the analysis – searching for a link between the timing of education policy reforms and cohort differences of equality of opportunity – is the one concerned with causes for changes of equality of opportunity and potentially indirectly with changes in the level of learning outcomes. That section of the analysis is more suggestive than conclusive, primarily due to the nature of the data, which makes it difficult to time cohort differences very precisely. Thus, the analyses do not seek to establish that the reforms identified have caused an observed change in an identified country, but rather to observe that similar reforms tend to be associated with similar between-cohort changes in equality of opportunity.

Longer-term equality of opportunity goes with higher learning outcomes

The analysis of the association between the levels of equality of opportunity and of proficiency on a systemic level shows that, in light of adult skills, equality of opportunity and the level of learning outcomes are positively associated for all parental education types, including the high type. The positive association is stronger for lower parental education types, which is to be expected when equality of opportunity is measured by the between-type difference in expected learning outcomes.

The positive association of equality of opportunity and the level of learning outcomes of the high parental education type is not insignificant, as equality indicators predict as much as half of the between-jurisdiction variation in learning outcomes. Notably, the positive association is stronger on more extended time frames, with longer-term equality indicators performing as well or better than short term ones in predicting proficiency in the short and very short terms. Also, models that predict proficiency using longer-term equality indicators do not seem to be improved by the inclusion of shorter-term equality indicators.

The finding that learning outcomes are positively associated with longer-term equality– income equality and equality of learning opportunity – is strengthened by the observation that equality indicators predict proficiency over time. Indeed, equality indicators from IALS perform nearly as well as indicators from both IALS and PIAAC in predicting proficiency in PIAAC. The augmented model using equality indicators from both surveys does generally achieve a better model fit, but also shows that the relative weight of predictive indicators (IALS in predicting PIAAC) increases when we progress from lower towards higher parental education types. This is in line with the observation that the positive association of proficiency with equality is, especially for the high parental education type, most apparent on the more extended time frames.

The results also imply that while income inequality and equality of opportunity are linked, they capture quite different aspects of equality. Either generally does quite well in predicting proficiency, but predictive power is improved by including the other in the predictive model. Their relationship also depends on parental education. The relative contribution of equality of opportunity to the predictive power of equality indicators is stronger for the lower parental education type, while the relative contribution of income inequality is higher for the higher parental education type.

The predictive power of the model seems to be reduced if the range of variation of equality of opportunity and especially income inequality is reduced. This is apparent in the strong influence of inclusion or exclusion of Chile and the United States on the ability of a model to predict proficiency using equality indicators. As shown above, a model constructed excluding Chile and United States is quite capable of predicting a large share of the variation of proficiency even when Chile and United States are included in the analysis, but that the predictive power of the model is much lower than for a full model constructed with the two countries in the analysis. The difference between the two models arises from the impact of income inequality on proficiency, which seems to be underestimated when the range of income inequality present in the sample is limited by the exclusion of countries with highly unequal income distributions. The eventual test for our interpretation will be how data from more unequal and less developed countries will fit the predictive model.

Together, the results have several practical and analytic implications.

Predicting proficiency with equality

First, our findings indicate that it is possible to predict proficiency over time using equality indicators, as higher proficiency would seem to be associated with longer-term equality. The element of time is

a crucial question in education policy because the timeframes involved in policy implementation are long, and the timeframes involved in the effects of policy becoming visible are much longer still. The lower the levels of the education system subjected to policy changes, the longer the timeframes over which the effects of the policy in adulthood are visible. Consequently, even if sufficient data has been collected to assess the effects of specific reforms, any such assessment comes very much after the policy change.

The situation can be ameliorated by the presence of advance indicators that can be used to assess the trajectory that the education system is on. The level of adult skills of a birth cohort can reasonably be expected to fall short of the previous cohort if they fall short of it at the end of compulsory education.

Equality of opportunity and income inequality would seem to be indicators that allow the prediction of learning outcomes over time. While improvements in equality are coincident with improvements in the learning outcomes of the groups with lower parental education, they also seem to predict the level of learning outcomes of the high parental education type quite well into the future.

In this respect, the results support using equality indicators as one source of information to anticipate later performance. They would seem to indicate that while, in the short term, losses in equality of opportunity may even be associated with improvements in the proficiency of the best-off groups, in the longer term it is notably difficult for even the best-off group to achieve high performance with very high inequality of opportunity or of income.

The ability to use equality indicators for advance scoping of the direction of the education system is not – and cannot be – a substitute for rigorous assessment of the effects of education policy reforms. Thus, it must be supplanted by a more detailed analysis of the treatment effects of reforms. However, as acquiring irrefutable proof of the effects of reforms is quite often a very long process indeed, tools for early diagnosis may be helpful, especially in cases where the education system seems to have taken a wrong turn, and the crucial indicators are showing worsening outcomes.

While hindsight provides a 20/20 vision that is not achievable going forward, we may note that, if the positive link between proficiency and equality had been better recognised, Finland might have placed more scrutiny on the results of national assessments in the early 2000s that suggested that socioeconomic differences were on the rise. There were signs that this might lead to a more general decline, especially after PISA results towards the end of the first decade of the millennium started to indicate that results were generally improving in countries where socioeconomic differences were decreasing and declining in those where the differences were increasing. Now, fifteen years after those warning signs, we can see that Finland has, in PISA, experienced a decline in both proficiency and equality that is not yet without parallel but is very notable.

However, it must be noted that we have not presented a general model for predicting proficiency over time, but simply analysed the association of equality of learning opportunity and proficiency across time and cohorts. As always, additional research on the causes that move education systems from high to low equality of opportunity and outcomes, or in the other direction, is evidently needed.

Opportunity seems to be opportunity

Our analyses have, in contrast to earlier research, separated different parental education types. The results indicate clearly that, especially over more extended time frames, there is no trade-off between equality of opportunity and the level of achievement, but rather a positive interdependence in which

the narrowing of differences between socioeconomic groups is associated with higher levels of achievement for all groups.

The results indicate that structural or policy factors that lead to high levels of equality of learning opportunity are also later associated with increased levels of proficiency for the high parental education type. This would suggest that policies that lead to increases in inequality of learning opportunity, and to temporary gains for the high parental education type, are likely to lead to decreased proficiency over the longer term. Here, the regular co-occurrence of higher equality of opportunity and a higher level of learning outcomes are in line with seminal literature on the equity-efficiency trade-off.

This implies that the opportunity hypothesis would seem to be a better predictor of the interplay of proficiency and equality of opportunity than the ability hypothesis. While the ability hypothesis treats the intergenerational association of social origins with learning outcomes as a consequence of revealed innate ability, the opportunity hypothesis treats it as a consequence of resources associated with social origins. Our observation that equality of opportunity is positively associated with proficiency for all parental education types is more consistent with the opportunity hypothesis.

This finding is not altogether surprising, as it is in line with the observation that heritability of educational achievement or of intelligence is still very much short of 100 %. While heritability should not be taken to be a measure of malleability, it is often expected to rise with general improvement and equalisation of environments, which would lead to the role of environmental variation in determining outcomes being diminished and the role of genetics being revealed as environment-induced variation is reduced.

While parental education is an ill-defined treatment, the results suggest that parental education is at least not entirely a genetic treatment, but that environmental inequalities are at least partially responsible for between-type differences in cognitive achievement. The results do not allow us to estimate whether or not it is partially genetic treatment, so that part of the between-type differences in cognitive achievement would persist if environments could be wholly equalised. At least part of the inequality of opportunity measured by standard methods of measuring between-type inequality seems to be inequality of circumstances, even when innate ability is not considered to be a circumstance.

These results would suggest, at least, that hypotheses which stipulate that inequalities of outcome associated with background directly match differences in innate ability – or even that outcome distributions have been narrowed beyond what would naturally emerge if environments were equal – would seem to be erroneous. Thus, even the developed and relatively egalitarian countries do not yet seem to be living in a genetic caste society, where between-type differences would be explained by between-type differences in innate ability.

This provides a cause for a reminder for education policy. Policy must recognise that not all individuals have the same innate abilities and dispositions, so as to avoid levelling down. However, thus far education systems do not seem to be so much in danger of suppressing the genetically gifted offspring of higher socioeconomic strata in the misguided attempt to impose equality of outcomes, but rather face the problem of failing to provide the individuals from lower socioeconomic backgrounds with equal opportunities. Countries are still be able to pursue the dual goals of excellence and equality, without the need to sacrifice one on the altar of the other.

Policy has an effect

The second part of the analyses, focused on the association of changes in equality of opportunity with changes in education policy, suggests that policy can make a difference. The rapidity of some of the changes in equality of opportunity between cohorts also implies that at least some of the opportunity gap is environmental, and not merely due to innate differences between socio-economic groups.

We observed above that cohort discontinuities are fairly, even surprisingly, stable from IALS to PIAAC, even for the groups that were youngest at the time of IALS, despite the 14-21 years separating the data collection of the two surveys. This implies that the experiences that each cohort has after its late teens exert relatively small effects on the level of equality of opportunity. Correspondingly, we may note that if the experiences of different cohorts have any significant effects on the level of equality of opportunity, the relevant experiences must take place in rather early stages of life. This introduces the possibility that equality of opportunity is affected by the education policy faced by the different cohorts.

To assess if this could be the case, we compared the picture painted by IALS and PIAAC on the changes in the level of equality of opportunity between cohorts. The primary analyses compared each group with an age group 10 years older – a comparison that, in ideal circumstances, would compare the last pre-reform cohorts to the first post-reform cohorts. IALS and PIAAC data continued to be highly convergent, even though measuring small changes with very limited sample sizes poses high demands on measurement accuracy.

After establishing that the prerequisites for associating cohort differences of equality of opportunity with education policy reforms were fulfilled, the analysis proceeded to observe how the cohort discontinuities are associated in time with education policy reforms. The research design and the rough cross-sectional data do not allow us to detect causality, but some patterns are observable. A large section of more significant discontinuities in equality of opportunity was associated with three types of reforms: changes in the length of compulsory schooling, comprehensive school reforms, and the reforms of the 1990s.

Across countries, we can observe that changes in the length of compulsory schooling are often associated with changes in equality of learning opportunity, with longer compulsory education associated with increased equality of learning opportunity. This result is consistent with earlier research on the socioeconomically differential effects of compulsory schooling, with lower socioeconomic strata deriving greater benefits from more extended compulsory education. Many countries also display marked discontinuities in association with comprehensive school reforms in the 1950s-1970s. These reforms seem, in line with earlier research, to be associated with increased levels of equality of learning opportunity. The same applies to reforms that have postponed tracking in the education system by other means. The third group of reforms are the reforms of the 1990s, many of which increased school choice and local autonomy, and seem to be associated with decreases in equality of opportunity.

Of interest to the general analysis is the observation that significant discontinuities associated with policy reforms are often followed by discontinuities in the opposite direction. In most cases, identifiable reforms are associated with increases in equality of learning opportunity, followed soon by decreases. This pattern may arise because low parental education types are, due to lower levels of economic and cultural resources, less cushioned against the effects of policy changes than high parental education types, whose higher resources mitigate the effect of environmental changes posed by the reforms.

The association of Finnish education policy reforms of the 1990s with decreased equality and proficiency have several implications for the interpretation of the trajectory of the Finnish education system. They would imply that the recently observed decline of Finland in PISA is part of a longer-term trend of declining proficiency and equality. It also raises the possibility that the apparent concurrent decline of proficiency and equality levels undergone by many mainly European countries might not only be interlinked but also linked to changes in the direction of education policy in the 1990s, rather than to impersonal cultural forces beyond the influence of policy. However, much more research is needed.

Overall, our results suggest that the pattern of equality differences between cohorts stays stable from early adulthood and that, in many cases, the discontinuities in equality correspond to changes in compulsory education. This would indicate that reforms of compulsory education have had a lasting effect on the populations and societies affected by them, while education policies after compulsory education have had a limited effect on the differences once established. From a policy perspective, this observation underlines the importance of compulsory education policy, as it shows the limits of higher levels of education in altering the basic pattern established early on.

13.3 Some paths for further research

Our results contribute to assessing whether education systems in developed countries face a trade-off between equality of opportunity and the level of learning outcomes, which would emerge if observed inequality of opportunity were merely to reflect between-group differences in innate ability. Thus, while the available data is very limited, the analyses contribute indirectly to the discussion on the relative contributions of environment and innate differences to between-group differences in cognitive outcomes. From a practical point of view, they also contribute to anticipating the evolution of the level of learning outcomes over time, by showing that inequality indicators can be used to predict the level of learning outcomes over time.

The analyses of the association of equality of learning opportunity and learning outcomes indicate that systems are still on the upward sloping part of the ability-opportunity curve. This result is consistent with existing literature on the determinants of learning outcomes, which implies that between-type differences in learning outcomes are only partially genetically mediated. Thus, even if the observed weak genetic explanation of education outcomes was in fact a strong genetic explanation as well, we should expect that all education systems would still benefit from increased equality of opportunity in the sense of smaller between-type differences in the expected value of learning outcomes. This suggests that overemphasised concern for the equity-efficiency trade-off is not only unnecessary but potentially harmful.

Trade-offs across time and socioeconomic groups

However, the results suggest that, for the high parental education type, there is a trade-off between relative national advantage in proficiency and the level of proficiency in international comparison, as well as between income inequality and the level of proficiency in international comparison. Thus, while the narrowing of the advantage of the high parental education type relative to the low parental education type is not achieved through lowered ability relative to the same group in other countries, the relative advantage in a national context is, by definition, narrowed. For strategy optimisation, it would be important to know how much value is attached to either relative performance. Is it better to

be smarter and more equal or a bit more ahead, even if the advantage would be paid for by a somewhat lower performance for oneself as well? How much individual variation is there in willingness to serve in heaven or reign in hell? How great is the externality cost of preserving relative national advantage if it is preserved by lowering performance?

These are all empirical and theoretical questions that go beyond the bounds of this work. However, it is good to note that, even in the absence of an opportunity trade-off in learning outcomes, it is not clear that all improvement in equality of opportunity is a no-cost policy. It is costly for the part of the population that benefits from existing inequality. This might very well be the case, especially if returns on lucky circumstances are large and/or very non-linear. If the distribution of circumstances is the factor that turns normally distributed ability inputs into power-law distributed wealth outcomes, then non-random luck may be a very precious commodity for those who possess it. And the more winner-take-all the market structure, the more it may be reasonable to gain a relative national advantage even by sacrificing performance.

Unlike previous research, our approach allows analysis over several timeframes, showing that the strength of the association between equality of learning opportunity and proficiency is also sensitive to the timeframe, especially for the high parental education type, for which the positive association of longer-term equality, of incomes or of opportunity, is most notable. This result suggests that the timeframe is also important when questions of trade-offs or incentives are posed regarding equality of outcomes or equality of learning opportunity in education. Depending on the length of foresight and the implicit discount rate, even investments that pay off over the long term may turn out not to make the cut in the short term. It may be assumed that this timeframe effect on the trade-offs involved would be combined with the effects of the intrinsic value of the relative national position. Higher parental education types would both pay for improved proficiency by losing the national advantage, and their improvement would take a longer time to be realised than the improvements for the lower parental education types.

Individuals tend to hold opinions that justify their rewards, even when the individual in question has been clearly cognisant of the fact that the outcome was predetermined and not due to individual merit. Thus, socioeconomic background should be associated with attitudes concerning issues that affect wealth or learning outcomes. For example, the attitude that success is due to individual ability and effort should be more common amongst those with higher socioeconomic backgrounds, but their attitudes should also be responsive to changes in the structure of market rewards. Increased winner-takes-all dynamics and increased income inequality, while they emphasise the role of luck for the outcomes, should be associated with increased emphasis on the role of innate ability in determining the outcomes. Recent empirical work would suggest that this is the case as, in more unequal countries, success is more often attributed to individual causes than in more equal ones. (Mijs, 2019)

Correspondingly, while the groups with high parental education in countries with lower levels of learning outcomes pay the highest price for the unequal distribution of luck, their attitudes should emphasise innate ability more than the attitudes of the corresponding group in countries with higher and more equal learning outcomes. Also, if the change in attitudes translates into policy and vice versa, attitudes and outcomes would form a virtuous/vicious cycle. In a virtuous circle, increased emphasis on environmental factors would translate into improved learning outcomes and reduced inequality of opportunity, which in turn would translate into attitudes that put more emphasis on environmental factors in determining outcomes. The opposite spiral would ensue if the expectation that success is largely determined by permanent and innate characteristics of the individual led to less effort being made to address differences in the environment. To analyse such a development, it might be possible to find data on the evolution of public perceptions on the relative contributions of innate

ability and social structures to cognitive outcomes. The association of inequality with the attribution of causes of success together with the first decreasing and then increasing income and wealth inequality in the 20th century would predict the general perception of what outcome differences are due to have evolved with the income and wealth distribution. Correspondingly, we might expect that public opinion on the causes of inequality of cognitive outcome emphasised environmental causes in the 1950s, 1960s and 1970s, when many countries experienced a narrowing of the socioeconomic achievement gap, while more essentialist attribution of outcomes to innate ability grew stronger in the 1990s and 2000s, when many countries experienced growing socioeconomic differences. Such long-term fluctuations would, quite possibly, fit with emerging theories of the cyclical development of social inequality and unrest. (Turchin, 2015, 2017)

Inequality and cognitive achievement

The ability of indicators of income inequality and equality of opportunity to predict cognitive performance over time also raises questions about the limits of this predictive ability. Here, on quite general measures of linguistic and quantitative reasoning, long-term income inequality predicted performance in all parental education groups well. Future research may show whether this ability extends to predicting performance in large-scale surveys of youth skills, like PISA, PIRLS and TIMMS, or to other measures of cognitive performance.

One of the possible areas in which to test the ability of income inequality to predict cognitive performance would be the area of intelligence testing. PIAAC and IALS domains are general enough and highly correlated, and we can expect them to load on *g*. This would suggest that income inequality should be negatively associated not only with levels of learning outcomes, but with observed intelligence as well.

One means of addressing this question would be to study the secular trends in income inequality and intelligence in different countries to see if trends in income inequality are matched by trends in intelligence. A glance at the identified anti-Flynn countries (Dutton et al., 2016) suggests that the question might merit more attention. Denmark, Norway and Finland have displayed signs of a negative Flynn effect since the 1990s. All of these countries exhibited a negative discontinuity in the 1990s in the direction of more inequality of opportunity, and have experienced rising income inequality in or since the 1990s. France has exhibited the anti-Flynn effect since the late 1990s, with income inequality increasing rapidly in the late 1990s and early 2000s. Britain has exhibited the negative Flynn effect since the mid- to late-1970s, and experienced rising income inequality since the very late 1970s.

These observations are of course suggestive at best, and there are counterexamples to the examples above. Netherlands and Estonia have shown signs of the anti-Flynn effect since the mid-1970s and early 2000s, respectively, and neither date matches a rise in income inequality. And of course, increasing income inequality has been much more prevalent in developed countries in the last decades than the anti-Flynn effect.

Thus, while the challenge for more thorough analyses will likely be the availability of representative intelligence data of good quality for a large enough sample of countries, further research is sorely needed. It may also be noted that analyses of the association between income inequality and cognitive performance should not be understood as a causal association with either acting as a cause of the other. Rather, it may be understood as work on better understanding what constitutes the phe-

nomenon to be explained. Causal understanding is then needed on the level of mechanisms to understand how the phenomenon is produced, and ill-defined causal analysis may provide a bridge between constitutive understanding of the phenomenon and causal understanding of its genesis.

Adult data and the skill effects of education policy reforms

The analysis of the skills effects of education policy reforms suffers from small sample sizes in subpopulations defined by both age and socioeconomic background. Despite the shortcomings, the results above showed that the between-type achievement gaps are relatively stable over time, as illustrated by the relatively good match between IALS and PIAAC.

The relative stability of at least certain aspects of the distribution of outcomes implies that upcoming cycles of PIAAC data collection will help provide better coverage of the relevant cohorts, as the observed cohorts will largely overlap with the first cycle of PIAAC. The increased sample size per birth cohort will hopefully allow future analyses to place equality of opportunities with greater precision by pooling the microdata. Here, analyses will be limited by the information available on whether the surveyed individuals have been affected by particular reforms or not. In cases where birth year provides an inexact proxy for affected and unaffected groups, collection of detailed contextual information on the schooling experiences would help in comparing the affected to the unaffected. The structure of the surveys poses clear limitations on the amount and nature of such contextual questions, but an adaptive survey would allow for more specific questions to be directed only to those individuals whose age is not sufficient to place them in the pre- or post-reform group.

Finding out what happened to Finland

Our results imply that something happened in Finland in the 1990s that led to a severe decline in equality of opportunity. Given the observed link between equality of opportunity and the level of learning outcomes, it is not inconceivable that the decline of results and of equality are a single phenomenon, and the origins of the change are to be found in the 1990s.

Our results are not sufficient to identify any of the changes of the 1990s as definite, or even likely, causes of the subsequent decline. They might, however, strengthen the case for a need to reassess the popular accounts of the causes of Finnish success in international large-scale surveys. Many of those accounts have made the connection between the reforms of the 1990s and the success of the early 2000s, and that account has become the standard account of the causes of Finnish success. These accounts may have limited the public discussion on the causes of the decline despite the fact that many of the experts have not privately subscribed to the popular public explanations. Our position as a superpower of education, based on success in large-scale surveys, has become part of the national myth, as have the putative causes of that success – autonomous master-level teachers, individuality and high levels of local autonomy – many of which only emerged in the Finnish schooling system in the 1990s. Reconsideration of what has happened has become difficult because it is at odds with how we have seen, and have wanted to see, ourselves.

As the learning outcomes in Finland have declined and grown increasingly unequal, public discussion has adapted very slowly to the changing situation. Emphasis has often been placed on the lack of statistical significance of short-term change, even while long-term change has become increasingly apparent, and on the limitations of individual surveys, most often PISA, when practically all sources have been consistent in showing that results are growing worse and increasingly unequal.

This evolution has been visible in national surveys of youth skills, in all international surveys of OECD and IEA (PISA, PIRLS, TIMSS), in the adult skills surveys, most notably PIAAC, as well as in administrative data on conscript skills. Thus, the observed trend is unlikely to be an artefact produced by the measurement instrument. We observe the same change with a multitude of instruments.

Different instruments are vulnerable to different problems when it comes to measurement, and consistency of the results across the instruments suggests that the results are not artefacts that are likely to appear in certain types of instruments. Thus, repeating surveys suffer from uncertainty as to how far the challenges of measurement invariance have been solved so that the results are effectively comparable over time. There is always a danger that the instrument undergoes changes in successive cycles that make the results increasingly incomparable over time and produce apparent trends that arise from changes in the instrument rather than from changes in reality. Also, with low stakes tests, there is always the possibility that changing attitudes affect test-taking motivation, and a rising or declining trend in the results does not reflect the change in actual skills but in the willingness of individuals to display their skills in a test. The first danger may be kept at bay by having several large-scale surveys, as surveys are unlikely to evolve in tandem so that all would produce the same artefactual trend.

The effect of test-taking motivation is more difficult to control for, especially if individuals cannot be trusted to provide accurate information on the level of effort they have expended in taking the test. Here the comparison of adult data with youth data may be helpful. Test-taking motivation may well affect results in adult surveys, but the necessary changes in attitudes are more specific, because a general change in attitudes would not produce a decline in results across cohorts. Thus, if a declining trend in learning outcomes – visible in both longitudinal youth data and cross-sectional adult data – across cohorts is due to test-taking motivation, there must be at least potentially observable differences in test-taking motivation between the cohorts. Most likely, there would need to be a process that affects test-taking motivation that would produce lower and lower motivation, which would at some age be crystallised and become a relatively permanent attribute of the cohort. Thus, adult data allows us to interrogate our interpretations from youth data. If loss of motivation co-occurs with loss of skills in one source but not in another, we may want to be cautious in giving it a very prominent role in any preliminary explanation.

Fin

This study has tried to approach fundamental questions on the possible conflict or complementarity between quality and equality in education. The issue is of general interest, as all societies must always face a tension between normative ideals and the limits that the material world poses on the realisation of those ideals. Here the inequality of the natural world is easily conscripted to support assertions that inequality is unavoidable in a society consisting mainly of naked apes. And indeed, it must be recognised that individuals differ in their innate ability, and it is very unlikely that all individuals could achieve the same cognitive outcomes even under optimal circumstances.

But the chain of reasoning does not usually stop here, with the observation of the importance of individual differences in innate ability in determining individual outcomes. Rather, the discussion easily turns into one that is as much, and more, about groups than it is about individuals. These groups of interest may be sexes, ethnic groups or socioeconomic groups, but the discussion is about whether the outcomes of the worse-off group are due to discrimination or to innate characteristics of the group.

This is not a new discussion, and various ideas about natural aristocracy have travelled with us through human history.

Now, it seems, essentialist explanations are making a serious comeback. In the areas of social, ethnic and sex differences, it has again become more common and accepted to attribute outcome differences to biological differences between groups rather than to environmental effects. It is important to recognise that humans are a species of animal and, as such, a special case of mammals, rather than immaterial cultural beings completely beyond biology. In this sense, the development has been very welcome. Unfortunately, much of the discussion is not so much about recognising the fundamental biological nature of humans but regressing to old ideas of natural aristocracy, where women, the poor, or those of different pigment are inherently inferior and thus quite deserving of their more modest positions in society. This is not so much a recognition of our biological nature as an attempt to use it to justify, rather than explain, differences between groups. We need to recognise the potential importance of genetics in understanding social selection and structures, but we should not neglect the fact that we are still very much unsure as to the origin of group differences in, e.g., cognitive achievement.

Our results would imply that equality of opportunity and excellence are not competing but complementary policy objectives. In showing this, the adult data would seem to reinforce the finding, arising from PISA, that small socioeconomic differences go together with high outcomes. These findings show that even developed countries can gain in both equalities of opportunity and proficiency. They imply that we do not have to be so worried about sacrificing excellence in our pursuit of (over-emphasised) equality. Rather, we need to become more worried about not mobilising the entire talent pool of society. The egalitarian education policies of the 1950s, 1960s and 1970s, often explicitly concerned with mobilising the untapped talents of lower socioeconomic strata, were often successful in their combined pursuit of equality and excellence. Thus, while the implications of the results are not limited to the effects of any specific policy and concern the general co-occurrence of equality of learning opportunity and level of learning outcomes, they do suggest that policies that lead to increases in the inequality of learning opportunity are unlikely to lead to increases in proficiency.

Any society that would give up some equality of opportunity to gain excellence would seem to lose a bit of both.

References

Abdellaoui, A., Hugh-Jones, D., Yengo, L. , Kemper, K. E., Nivard, M. G., Veul, L., . . . Wray, N. J. (2019) Genetic correlates of social stratification in Great Britain. *Nature Human Behaviour,* 3, 1332–1342. doi:10.1038/s41562-019-0757-5

Aho, E., Pitkänen, K., Sahlberg, P. (2006). *Policy development and reform principles of basic and secondary education in Finland since 1968 (English).* (Education working paper series; no. 2). Washington, DC: World Bank. http://documents.worldbank.org/curated/en/124381468038093074/Policy-development-and-reform-principles-of-basic-and-secondary-education-in-Finland-since-1968

Allegrini, A. G., Selzam, S., Rimfeld, K., von Stumm, S., Pingault, J. B., & Plomin, R. (2019). Genomic prediction of cognitive traits in childhood and adolescence. *Molecular psychiatry*, *24*(6), 819. doi:10.1038/s41380-019-0394-4

Allegrini, A. G., Karhunen, V., Coleman, J. R., Selzam, S., Rimfeld, K., von Stumm, S., ... & Plomin, R. (2019). Multivariable GE interplay in the prediction of educational achievement. *bioRxiv*, 865360.

Altinok, N., Diebolt, C., & Demeulemeester, J. L. (2014). A new international database on education quality: 1965–2010. *Applied Economics*, *46*(11), 1212-1247.

Arneson, R. J. (1990). Liberalism, distributive subjectivism, and equal opportunity for welfare. *Philosophy & public affairs*, 158-194.

Arneson, R. J. (1999). Against Rawlsian equality of opportunity. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *93*(1), 77-112.

Arneson, R. J. (2000). Perfectionism and politics. *Ethics*, *111*(1), 37-63.

Arrow, K. J. (1973). Higher education as a filter. *Journal of public economics*, *2*(3), 193-216.

Asbury, K., & Plomin, R. (2013). *G is for genes: The impact of genetics on education and achievement* (Vol. 24). John Wiley & Sons.

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of economic theory*, *2*(3), 244-263.

Atkinson, A. B. (2015). *Inequality*. Harvard University Press.

Autor, D. H., Katz, L. F., & Kearney, M. S. (2008). Trends in US wage inequality: Revising the revisionists. *The Review of economics and statistics*, *90*(2), 300-323.

Ayorech, Z., Krapohl, E., Plomin, R., & von Stumm, S. (2017). Genetic influence on intergenerational educational attainment. *Psychological science*, *28*(9), 1302-1310.

Baker, D. P., Eslinger, P., Benavides, M., Peters, E., Dieckmann, N. F., & Leon, J. (2015). The cognitive impact of the education revolution: A possible cause of the Flynn Effect on population IQ. *Intelligence*, *49*, 144-158.

Banks, J., & Mazzonna, F. (2012). The effect of education on old age cognitive abilities: evidence from a regression discontinuity design. *The Economic Journal*, *122*(560), 418-448. doi:10.1111/j.1468-0297.2012.02499.x

Barro, R. J., & Lee, J. W. (2015). *Education matters. Global schooling gains from the 19th to the 21st century*. Oxford University Press.

Barro, R. J., & Lee, J. W. (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of development economics*, *104*, 184-198.

Barry, B. (1988). Equal Opportunity and Moral Arbitrariness. In N. E. Bowie (Ed.), *Equal opportunity* (pp. 23-44). Boulder: Westview Press.

Barry, B. (1989). *Theories of justice: a treatise on social justice* (Vol. 16). University of California Press.

Barry, B. (1990). *Political argument: a reissue with new introduction* (No. 20). University of California Press.

Beebee, H., Hitchcock, C., & Price, H. (Eds.). (2017). *Making a difference: Essays on the philosophy of causation*. Oxford University Press.

Belsky, D. W., Domingue, B. W., Wedow, R., Arseneault, L., Boardman, J. D., Caspi, A., ... & Moffitt, T. E. (2018). Genetic analysis of social-class mobility in five longitudinal studies. *Proceedings of the National Academy of Sciences*, *115*(31), E7275-E7284.

Berg, A., & Ostry, J. D. (2011). Equality and efficiency. *Finance & Development*, *48*(3), 12-15.

Berg, A., Ostry, J. D., & Zettelmeyer, J. (2012). What makes growth sustained?. *Journal of Development Economics*, *98*(2), 149-166.

Berg, J. J., Harpak, A., Sinnott-Armstrong, N., Joergensen, A. M., Mostafavi, H., Field, Y., ... & Coop, G. (2019). Reduced signal for polygenic adaptation of height in UK Biobank. *ELife*, *8*, e39725.

Berlin, I. (1958). Two Concepts of Liberty, reprinted in Isaiah Berlin (1982), Four Essays on Liberty. In: Oxford, Oxford University Press.

Bernelius, V. (2008). *Lähi (ö) koulu: Helsingin koulut ja kaupunginosat EU: n Urban II-ohjelman kouluhankkeissa*. Helsingin kaupungin tietokeskus.

Bernelius, V. (2013). Eriytyvät kaupunkikoulut: Helsingin peruskoulujen oppilaspohjan erot, perheiden kouluvalinnat ja oppimistuloksiin liittyvät aluevaikutukset osana kaupungin eriytymiskehitystä.

Björklund, A., Jäntti, M., & Roemer, J. E. (2012). Equality of opportunity and the distribution of long-run income in Sweden. *Social choice and welfare*, *39*(2-3), 675-696.

Björklund, A., Jäntti, M., & Solon, G. (2007). Nature and nurture in the intergenerational transmission of socioeconomic status: Evidence from Swedish children and their biological and rearing parents. *The BE Journal of Economic Analysis & Policy*, *7*(2). Retrieved from http://papers.nber.org/papers/w12985

Bobbio, N. (1996). *Left and right: The significance of a political distinction*. University of Chicago Press.

Böckerman, P., Hämäläinen, U., & Uusitalo, R. (2009). Labour market effects of the polytechnic education reform: The Finnish experience. *Economics of Education Review*, *28*(6), 672-681.

Böhlmark, A., & Lindahl, M. (2012). *Har den växande friskolesektorn varit bra för elevernas utbildningsresultat på kort och lång sikt?*. Institutet för arbetsmarknads-och utbildningspolitisk utvärdering (IFAU).

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of educational research*, *73*(2), 125-230.

Bourguignon, F., Ferreira, F. H., & Menéndez, M. (2007). Inequality of opportunity in Brazil. *Review of income and Wealth*, *53*(4), 585-618.

Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An expanded view of complex traits: from polygenic to omnigenic. *Cell*, *169*(7), 1177-1186.

Branigan, A. R., McCallum, K. J., & Freese, J. (2013). Variation in the heritability of educational attainment: An international meta-analysis. *Social forces*, *92*(1), 109-140.

Bratsberg, B., & Rogeberg, O. (2018). Flynn effect and its reversal are both environmentally caused. *Proceedings of the National Academy of Sciences*, *115*(26), 6674-6678.

Bratti, M., Checchi, D., & De Blasio, G. (2008). Does the expansion of higher education increase the equality of educational opportunities? Evidence from Italy. *Labour*, *22*, 53-88.

Breen, R., & Goldthorpe, J. H. (1999). Class inequality and meritocracy: a critique of Saunders and an alternative analysis1. *The British journal of sociology*, *50*(1), 1-27.

Broer, M., Bai, Y., & Fonseca, F. (2019). *Socioeconomic Inequality and Educational Outcomes: Evidence from Twenty Years of TIMSS* (Vol. 5). Springer.

Brunello, G., & Checchi, D. (2007). Does school tracking affect equality of opportunity? New international evidence. *Economic policy*, *22*(52), 782-861.

Bukodi, E., & Goldthorpe, J. (2011). Social class returns to higher education: chances of access to the professional and managerial salariat for men in three British birth cohorts. *Longitudinal and Life Course Studies*, *2*(2), 185-201.

Bukodi, E., & Goldthorpe, J. H. (2012). Decomposing 'social origins': The effects of parents' class, status, and education on the educational attainment of their children. *European sociological review*, *29*(5), 1024-1039. doi:10.1093/esr/jcs079

Bukodi, E., & Goldthorpe, J. H. (2016). Educational attainment-relative or absolute-as a mediator of intergenerational class mobility in Britain. *Research in Social Stratification and Mobility*, *43*, 5-15.

Cameron, S., Daga, R., & Outhred, R. (2018). Setting out a conceptual framework for measuring equity in learning. In *Handbook of Measuring Equity in Education*. Montreal, Quebec, Canada: UNESCO Institute for Statistics.

Caplan, B. (2018). *The case against education: Why the education system is a waste of time and money*. Princeton University Press.

Cattell, R. B. (1950). The fate of national intelligence: Test of a thirteen-year prediction. *The Eugenics Review*, *42*(3), 136.

Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental psychology*, *27*(5), 703.

Chabris, C. F., Lee, J. J., Cesarini, D., Benjamin, D. J., & Laibson, D. I. (2015). The fourth law of behavior genetics. *Current Directions in Psychological Science*, *24*(4), 304-312.

Checchi, D. & Flabbi, L. (2007). Intergenerational Mobility and Schooling Decisions in Germany and Italy: The Impact of Secondary School Tracks. *IDEAS Working Paper Series from RePEc*.

Checchi, D., Ichino, A., & Rustichini, A. (1999). More equal but less mobile?: Education financing and intergenerational mobility in Italy and in the US. *Journal of public economics*, *74*(3), 351-393.

Checchi, D., & Peragine, V. (2010). Inequality of opportunity in Italy. *The Journal of Economic Inequality*, *8*(4), 429-450.

Checchi, D., Peragine, V., & Serlenga, L. (2010). *Fair and unfair income inequalities in Europe*. (Society for the Study of Economic Inequality (ECINEQ) Working Papers Vol. 174). ECINEQ, Society for the Study of Economic Inequality.

Chetty, R., Hendren, N., Jones, M. R., & Porter, S. R. (2018). *Race and economic opportunity in the United States: An intergenerational perspective* (No. w24441). National Bureau of Economic Research.

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of opportunity? The geography of intergenerational mobility in the United States. *The Quarterly Journal of Economics*, *129*(4), 1553-1623.

Chetty, R., Hendren, N., Kline, P., Saez, E., & Turner, N. (2014). Is the United States still a land of opportunity? Recent trends in intergenerational mobility. *American Economic Review*, *104*(5), 141-47.

Chmielewski, A. K. (2019). The Global Increase in the Socioeconomic Achievement Gap, 1964 to 2015. *American Sociological Review*, *84(3)*, 517-544. doi:10.1177/0003122419847165

Cingano, F. (2014). *Trends in Income Inequality and its Impact on Economic Growth.* (OECD Social, Employment and Migration Working Papers, No. 163). OECD Publishing, Paris.

Clark, G. (2015). *The son also rises: Surnames and the history of social mobility* (Vol. 49). Princeton University Press.

Cohen, G. A. (1995). *Self-ownership, freedom, and equality*. Cambridge University Press.

Cohen, G. A. (2009). *Rescuing justice and equality*. Harvard University Press.

Cohen, G. A. (2011). *On the currency of egalitarian justice, and other essays in political philosophy*. Princeton University Press.

Cole, S. R., & Frangakis, C. E. (2009). The consistency statement in causal inference: a definition or an assumption?. *Epidemiology*, *20*(1), 3-5. doi:10.1097/EDE.0b013e31818ef366

Coleman, J. S. (1966). *Equality of educational opportunity.* U.S. Dept. of Health, Education, and Welfare, Office of Education, National Center for Education Statistics.

Colman, A. M. (2016). Race differences in IQ: Hans Eysenck's contribution to the debate in the light of subsequent research. *Personality and Individual Differences*, *103*, 182-189.

Colodro-Conde, L., Rijsdijk, F., Tornero-Gómez, M. J., Sánchez-Romera, J. F., & Ordoñana, J. R. (2015). Equality in educational policy and the heritability of educational attainment. *PLoS One*, *10*(11). doi:10.1371/journal.pone.0143796

Constant, B. (2010). *De la liberté des anciens comparée à celle des modernes*. Fayard/Mille et une nuits.

Corak, M. (Ed.). (2004). *Generational income mobility in North America and Europe*. Cambridge University Press.

Corak, M. (2013). Income inequality, equality of opportunity, and intergenerational mobility. *Journal of Economic Perspectives*, *27*(3), 79-102.

Economic Policy Council. (2018). Economic Policy Council Report 2017. VATT Institute for Economic Research. ISBN 978-952-274-210-0

Dandine-Roulland, C., Bellenguez, C., Debette, S., Amouyel, P., Génin, E., & Perdry, H. (2016). Accuracy of heritability estimations in presence of hidden population stratification. *Scientific reports*, *6*. doi:10.1038/srep26471

de-Shalit, A., & Wolff, J. (2011). The Apparent Asymmetry of Responsibility. In C. Knight & Z. Stemplowska (Eds.), *Responsibility and distributive justice*. Oxford University Press.

Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, *210*, 2-21. doi:https://doi.org/10.1016/j.socscimed.2017.12.005

Delannoy, F. Education Reforms in Chile, 1980-98: A Lesson in Pragmatism. (Country Studies: Education Reform and Management Publication.) Washington, DC: World Bank.

Derex, M., Bonnefon, J. F., Boyd, R., & Mesoudi, A. (2019). Causal understanding is not necessary for the improvement of culturally evolving technology. *Nature human behaviour*, *3*(5), 446. doi:10.1038/s41562-019-0567-9

Dickens, W. T., & Flynn, J. R. (2001). Heritability estimates versus large environmental effects: the IQ paradox resolved. *Psychological review*, *108*(2), 346.

Donati, G., Dumontheil, I., & Meaburn, E. L. (2019). Genome-Wide Association Study of Latent Cognitive Measures in Adolescence: Genetic Overlap With Intelligence and Education. *Mind, Brain, and Education*. doi:10.1111/mbe.12198

Duncan, L., Shen, H., Gelaye, B., Meijsen, J., Ressler, K., Feldman, M., ... & Domingue, B. (2019). Analysis of polygenic risk score usage and performance in diverse human populations. *Nature communications*, *10*(1), 1-9.

Durkheim, E. (1893). *De la division du travail social*. Les classiques des sciences sociales. http://classiques.uqac.ca/

Dutton, E., & Lynn, R. (2013). A negative Flynn effect in Finland, 1997–2009. *Intelligence*, *41*(6), 817-820.

Dutton, E., van der Linden, D., & Lynn, R. (2016). The negative Flynn Effect: A systematic literature review. *Intelligence*, *59*, 163-169.

Dworkin, R. (1981). What is equality? Part 1: Equality of welfare. *Philosophy & public affairs*, 185-246.

Dworkin, R. (2002). *Sovereign virtue: The theory and practice of equality*. Harvard university press.

Dworkin, R. (2011). *Justice for hedgehogs*. Harvard University Press.

Engzell, P., & Tropf, F. C. (2019). Heritability of education rises with intergenerational mobility. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1912998116

Ericsson, A., & Pool, R. (2016). *Peak: Secrets from the new science of expertise*. Houghton Mifflin Harcourt.

Eysenck, H. (1998). Intelligence: A new look: Routledge.

Eysenck, H. J. (1973). *The inequality of man* (Vol. 3). London: Temple Smith.

Eysenck, H. J. (1979). *The measurement of intelligence*. Springer Science & Business Media.

Fan, Y., Yi, J., & Zhang, J. (2015). *The Great Gatsby Curve in China: Cross-Sectional Inequality and Intergenerational Mobility*. Working Paper, CUHK, Hongkong.

Ferreira, F. H., & Gignoux, J. (2013). The measurement of educational inequality: Achievement and opportunity. *The World Bank Economic Review*, *28*(2), 210-246.

Ferreira, F. H., & Gignoux, J. (2008). *The measurement of inequality of opportunity: Theory and an application to Latin America*. The World Bank.

Ferreira, F. H., & Peragine, V. (2015). *Equality of opportunity: Theory and evidence*. The World Bank.

Figlio, D. N., Freese, J., Karbownik, K., & Roth, J. (2017). Socioeconomic status and genetic influences on cognitive development. *Proceedings of the National Academy of Sciences*, *114*(51), 13441-13446. doi: 10.1073/pnas.1708491114

Fischbein, S. (1980). IQ and social class. *Intelligence*, *4*(1), 51-63.

Flabbi, L. (1999). *Returns to Schooling in Italy, OLS, IV and Gender Differences*. (Working Paper: Serie di econometria ed economia applicata, N. 1, 1999) Università Bocconi.

Fleurbaey, M. (2008). *Fairness, responsibility, and welfare*. Oxford University Press.

Fleurbaey, M., & Peragine, V. (2013). Ex ante versus ex post equality of opportunity. *Economica*, *80*(317), 118-130. doi:10.1111/j.1468-0335.2012.00941.x

Fleurbaey, M., & Schokkaert, E. (2009). Unfair inequalities in health and health care. *Journal of health economics*, *28*(1), 73-90.

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological bulletin*, *95*(1), 29.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological bulletin*, *101*(2), 171.

Flynn, J. R. (2007). *What is intelligence?: Beyond the Flynn effect*. Cambridge University Press.

Flynn, J. R. (2012). *Are we getting smarter?: Rising IQ in the twenty-first century*. Cambridge University Press.

Fort, M. (2006). Educational reforms across Europe: A toolbox for empirical research. unpublished paper.

Fort, M., Schneeweis, N., & Winter-Ebmer, R. (2014). *More Schooling, More Children? Compulsory Schooling and Fertility in Europe* (No. 5068). CESifo Group Munich.

Frankfurt, H. G. (2015). *On inequality*. Princeton University Press.

Frieden, T. R. (2017). Evidence for health decision making—beyond randomized, controlled trials. *New England Journal of Medicine*, *377*(5), 465-475. doi:10.1056/NEJMra1614394

Gamoran, A., & Mare, R. D. (1989). Secondary school tracking and educational inequality: Compensation, reinforcement, or neutrality?. *American journal of Sociology*, *94*(5), 1146-1183.

Garrouste, C. (2010). *100 years of educational reforms in Europe: a contextual database* (MPRA Paper No. 31853). University Library of Munich, Germany.

Glass, T. A., Goodman, S. N., Hernán, M. A., & Samet, J. M. (2013). Causal inference in public health. *Annual review of public health*, *34*, 61-75.

Goodman, M. J., Sands, A. M., & Coley, R. J. (2015). America's Skills Challenge: Millennials and the Future. *Educational Testing Service*.

Gottfredson, L. S. (2000). Equal potential: A collective fraud. *Society*, *37*(5), 19. doi:10.1007/s12115-000-1033-y

Gottfredson, L. S. (2004). Intelligence: is it the epidemiologists' elusive" fundamental cause" of social class inequalities in health?. *Journal of personality and social psychology*, *86*(1), 174.

Gottfredson, L. S. (2005). What if the hereditarian hypothesis is true?. *Psychology, Public Policy, and Law*, *11*(2), 311-319.

Gottfredson, L. S. (2011). Intelligence and social inequality: Why the biological link. *The Wiley-Blackwell handbook of individual differences. Blackwell Publishing Ltd*, 538-75.

Gottfredson, L. S. (2016). Hans Eysenck's theory of intelligence, and what it reveals about him. *Personality and Individual Differences*, *103*, 116-127. doi:10.1016/j.paid.2016.04.036

Green, A., Green, F., & Pensiero, N. (2015). Cross-country variation in adult skills inequality: Why are skill levels and opportunities so unequal in Anglophone countries?. *Comparative Education Review*, *59*(4), 595-618. doi:10.1086/683101

Green, S. J. (1989). Emile Durkheim on human talents and two traditions of social justice. *British Journal of Sociology*, 97-117.

Griffiths, P. E. (2002). What is innateness?. *The Monist*, *85*(1), 70-85.

Grömping, U. (2015). Variable importance in regression models. *Wiley Interdisciplinary Reviews: Computational Statistics*, *7*(2), 137-152.

Gustafsson, J. E. (2008). Schooling and intelligence: Effects of track of study on level and profile of cognitive abilities. *Extending intelligence: Enhancement and new constructs*, 37-59.

Haier, R. J. (2016). *The neuroscience of intelligence*. Cambridge University Press.

Halpern, D. F. (2000). *Sex differences in cognitive abilities*. Psychology press.

Hansen, K. Y., Gustafsson, J. E., Rosén, M., Sulkunen, S., Nissinen, K., Kupari, P., ... & Mejding, J. (2014). *Northern Lights on TIMSS and PIRLS 2011: Differences and similarities in the Nordic countries*. Nordic Council of Ministers.

Hanushek, E. A., & Woessmann, L. (2015). *The knowledge capital of nations: Education and the economics of growth*. MIT press.

Härnqvist, K. (1968). Relative changes in intelligence from 13 to 18: Results. *Scandinavian journal of psychology*, *9*(1), 65-82.

Hauspie, R. C., Vercauteren, M., & Susanne, C. (1997). Secular changes in growth and maturation: an update. *Acta Paediatrica*, *86*(S423), 20-27. doi:10.1111/j.1651-2227.1997.tb18364.x

Haworth, C. M., Wright, M. J., Luciano, M., Martin, N. G., de Geus, E. J., van Beijsterveldt, C. E., ... & Kovas, Y. (2010). The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Molecular psychiatry*, *15*(11), 1112-1120. doi:10.1038/mp.2009.55

Hayek, F. A. von. (2013). *The constitution of liberty: The definitive edition*. Routledge.

Hayek, F. A. von. (2012). *Law, legislation and liberty: a new statement of the liberal principles of justice and political economy*. Routledge.

Hernán, M. A. (2018). The C-word: scientific euphemisms do not improve causal inference from observational data. *American journal of public health*, *108*(5), 616-619. doi:10.2105/AJPH.2018.304337

Hernán, M. A., Hernández-Díaz, S., & Robins, J. M. (2013). Randomized trials analyzed as observational studies. *Annals of internal medicine*, *159*(8), 560-562. doi:10.7326/0003-4819-159-8-201310150-00709

Hernán, M. A., & Robins, J. M. (2020). *Causal inference*. Boca Raton: Chapman & Hall/CRC.

Hernán, M. A., & Taubman, S. L. (2008). Does obesity shorten life? The importance of well-defined interventions to answer causal questions. *International journal of obesity*, *32*(S3), S8-S14. doi:10.1038/ijo.2008.82

Herrnstein, R. J. (1971). IQ. *Atlantic Monthly*. September. 43-64.

Herrnstein, R. J. (1973). *IQ in the meritocracy*. Little Brown.

Hill, W. D., Marioni, R. E., Maghzian, O., Ritchie, S. J., Hagenaars, S. P., McIntosh, A. M., ... & Deary, I. J. (2019). A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. *Molecular psychiatry*, *24*(2), 169.

Hill, W. D., Davies, N. M., Ritchie, S. J., Skene, N. G., Bryois, J., Bell, S., ... & Liewald, D. C. (2019). Genome-wide analysis identifies molecular systems and 149 genetic loci associated with income. *Nature Communications*, *10*(1), 1-16.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association*, *81*(396), 945-960.

Hur, Y. M., & Bates, T. (2019). Genetic and Environmental Influences on Cognitive Abilities in Extreme Poverty. *Twin Research and Human Genetics*, *22*(5), 297-301. doi:10.1017/thg.2019.92

Huxley, A. (2008). *Brave new world*. Vintage Digital.

Hyytinen, A., Ilmakunnas, P., Johansson, E., & Toivanen, O. (2019). Heritability of lifetime earnings. *The Journal of Economic Inequality*, 17(3), 319-335. doi:10.1007/s10888-019-09413-x

Jacob, W. J., & Holsinger, D. B. (2008). Inequality in education: A critical analysis. In *Inequality in education* (pp. 1-33). Springer, Dordrecht.

Jacobs, L. A. (2004). *Pursuing equal opportunities: the theory and practice of egalitarian justice*. Cambridge University Press.

Jæger, M. M. (2007). Educational mobility across three generations: The changing impact of parental social class, economic, cultural and social capital. *European Societies*, *9*(4), 527-550.

James, W. (1882). *Rationality, activity and faith*. University of Michigan: Humanities Text Initiative.

Jensen, A. (1969). How much can we boost IQ and scholastic achievement. *Harvard educational review*, *39*(1), 1-123.

Jensen, A. (1973). *Educability and Group Differences*. Harper & Row.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport (Conn.): Praeger.

Jerrim, J., & Macmillan, L. (2015). Income inequality, intergenerational mobility, and the Great Gatsby Curve: Is education the key?. *Social Forces*, *94*(2), 505-533.

Jerskey, B. A., Panizzon, M. S., Jacobson, K. C., Neale, M. C., Grant, M. D., Schultz, M., ... & Lyons, M. J. (2010). Marriage and divorce: A genetic perspective. *Personality and individual differences*, *49*(5), 473-478. doi:10.1016/j.paid.2010.05.007

Kanaya, T., & Ceci, S. J. (2011). The Flynn effect in the WISC subtests among school children tested for special education services. *Journal of Psychoeducational Assessment*, *29*(2), 125-136. doi:10.1177/0734282910370139

Karhunen, H., & Uusitalo, R. (2017). 50 vuotta koulutusmahdollisuuksien tasa-arvoa. *Yhteiskuntapolitiikka*, 82(3), 296-303.

Kaufman, J. S. (2019). Commentary: Causal inference for social exposures. *Annual review of public health*, *40*, 7-21. doi:10.1146/annurev-publhealth-040218-043735

Kelsen, H. (1967). *Pure theory of law*. University of California Press.

Kelsen, H. (1945). *General theory of law and state*. Harvard University Press.

Kendler, K. S., Turkheimer, E., Ohlsson, H., Sundquist, J., & Sundquist, K. (2015). Family environment and the malleability of cognitive ability: A Swedish national home-reared and adopted-away cosibling control study. *Proceedings of the National Academy of Sciences*, *112*(15), 4612-4617.

Kenworthy, L. (1995). Equality and efficiency: The illusory tradeoff. *European Journal of Political Research*, *27*(2), 225-254.

Kerminen, S., Martin, A. R., Koskela, J., Ruotsalainen, S. E., Havulinna, A. S., Surakka, I., ... & Ripatti, S. (2019). Geographic Variation and Bias in the Polygenic Scores of Complex Diseases and Traits in Finland. *The American Journal of Human Genetics*.

Kevles, Daniel J. *In the name of eugenics: Genetics and the uses of human heredity*. No. 95. Harvard University Press, 1995.

Khalidi, M. A. (2007). Innate cognitive capacities. *Mind & Language*, *22*(1), 92-115.

Kirsch, I. S. (2001). The international adult literacy survey (IALS): Understanding what was measured. *ETS Research Report Series*, *2001*(2), i-61.

Kodelja, Z. (2016). Equality of opportunity and equality of outcome. *CEPS Journal*, *6*(2), 9-24.

Koivunen, S. (2007). Suomalaismiesten kognitiivisen kykyprofiilin muutokset 1988-2001: Flynnin efektiä suomalaisessa aineistossa? Master's thesis. Jyväskylän yliopisto.

Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsson, B. J., Young, A. I., Thorgeirsson, T. E., ... & Gudbjartsson, D. F. (2018). The nature of nurture: Effects of parental genotypes. *Science*, *359*(6374), 424-428.

Korsgaard, C. M., Cohen, G. A., Geuss, R., Nagel, T., & Williams, B. (1996). *The sources of normativity*. Cambridge University Press.

Kosunen, S. (2016). *Families and the social space of school choice in urban Finland*. (Studies in Educational Sciences 267). University of Helsinki, Institute of Behavioural Sciences.

Kraaykamp, G., & Van Eijck, K. (2010). The intergenerational reproduction of cultural capital: A threefold perspective. *Social forces*, *89*(1), 209-231.

Krapohl, E., & Plomin, R. (2016). Genetic link between family socioeconomic status and children's educational achievement estimated from genome-wide SNPs. *Molecular psychiatry*, *21*(3), 437.

Kuznets, S. (1955). Economic growth and income inequality. *The American economic review*, *45*(1), 1-28.

Lefranc, A., Pistolesi, N., & Trannoy, A. (2009). Equality of opportunity and luck: Definitions and testable conditions, with an application to income in France. *Journal of Public Economics*, *93*(11-12), 1189-1207.

Lem, S. (2014). *Solaris*. Pro Auctore Wojciech Zemek.

Levin, B. (2003). Approaches to equity in policy for lifelong learning. *Education and Training Policy Division, OECD.*

Lewontin, R. (2010). *Biology as ideology: The doctrine of DNA*. House of Anansi Press.

Lewontin, R. C. (1970). Race and intelligence. *Bulletin of the Atomic Scientists*, *26*(3), 2-8.

Lippert-Rasmussen, K. (2011). Luck egalitarianism and group responsibility. *Responsibility and distributive justice*, 98-114.

Lynn, R., Vanhanen, T., & Stuart, M. (2002). *IQ and the wealth of nations*. Greenwood Publishing Group.

Mandeville, B. (1989). *The fable of the bees*. Penguin.

Marioni, R. E., Davies, G., Hayward, C., Liewald, D., Kerr, S. M., Campbell, A., ... & Hastie, N. D. (2014). Molecular genetic contributions to socioeconomic status and intelligence. *Intelligence*, *44*, 26-32. doi:10.1016/j.intell.2014.02.006

Marks, G. N. (2013). *Education, social background and cognitive ability: The decline of the social*. Routledge.

Marrero, G. A., & Rodríguez, J. G. (2013). Inequality of opportunity and growth. *Journal of Development Economics*, *104*, 107-122.

Maurin, E., & McNally, S. (2007). Educational Effects of Widening Access to the Academic Track: A Natural Experiment. CEE DP 85. *Centre for the Economics of Education (NJ1)*.

Maxwell, J. (1954). Intelligence, fertility and the future: A report on the 1947 Scottish Mental Survey. *Eugenics Quarterly*, *1*(4), 244-247.

Meghir, C., & Palme, M. (2005). Educational reform, ability, and family background. *American Economic Review*, *95*(1), 414-424.

Mijs, J. J. (2019). The paradox of inequality: income inequality and belief in meritocracy go hand in hand. *Socio-Economic Review*.

Lin, M. J. (2017). *The intergenerational transmission of educational attainment revisited: the effects of socioeconomic background, genetic inheritance, and cohort.* Master's thesis. The University of North Carolina at Chapel Hill).

Meschi, E., & Scervini, F. (2014). A new dataset on educational inequality. *Empirical Economics*, *47*(2), 695-716.

Milanovic, B. (2019). All the Ginis Database. (version February 2019). Retrieved from: https://www.gc.cuny.edu/Page-Elements/Academics-Research-Centers-Initiatives/Centers-and-Institutes/Luxembourg-Income-Study-Center/Branko-Milanovic,-Senior-Scholar/Datasets

Mises, L. von. (1951). *Socialism: An Economic and Sociological Analysis*. Yale University Press.

Mises, L. von. (2006). *Human action: A treatise on economics.* Liberty Fund.

Montalvan, P., & Lemay, M. (2013). Chapter 10: Field Operations. In *Technical Report of the Survey of Adult Skills (PIAAC)*. Organisation for Economic Development and Co-operation OECD.

Herrnstein, R. J., & Murray, C. (2010). *The bell curve: Intelligence and class structure in American life*. Simon and Schuster.

Murray, T. S., Kirsch, I. S., & Jenkins, L. B. (1998). *Adult Literacy in OECD Countries: Technical Report on the First International Adult Literacy Survey*. US Government Printing Office. Washington, DC.

Nagel, T. (2012). *Mortal questions*. Cambridge University Press.

Natkhov, T., & Kozina, N. (2012). Inequality of Educational Opportunity in a Cross-Section of Countries: Empirical Analysis of 2009 PISA Data. *Higher School of Economics Research Paper No. BRP*, *7*.

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Group differences in IQ are best understood as environmental in origin. *The American psychologist*, *67*(6), 503-504.

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: new findings and theoretical developments. *American psychologist*, *67*(2), 130.

Nozick, R. (1974). *Anarchy, state, and utopia*. Basic Books.

OECD. (2000). *Literacy in the information age: Final report of the International Adult Literacy Survey*. OECD Publishing. Organisation for Economic Co-operation and Development (OECD).

OECD. (2004). *Learning for Tomorrow's World: First results from PISA 2003*. OECD Publishing. Organisation for Economic Co-operation and Development (OECD).

OECD. (2007). *PISA 2006: Science Competencies for Tomorrow's World: Volume 1: Analysis*. OECD Publishing. Organisation for Economic Co-Operation and Development (OECD).

OECD. (2010). *PISA 2009 results: overcoming social background: equity in learning opportunities and outcomes (volume II)*. OECD Publishing. Organisation for Economic Co-Operation and Development (OECD).

OECD. (2013a). *PISA 2012 results: excellence through equity: giving every student the chance to succeed (volume II).* OECD Publishing. Organisation for Economic Co-Operation and Development (OECD).

OECD. (2013b). *Skills Outlook 2013 - First Results from the Survey of Adult Skills.* OECD Publishing. Organisation for Economic Co-Operation and Development (OECD).

OECD. (2013c). *Technical Report of the Survey of Adult Skills (PIAAC) (Pre-publication copy 17 October 2013)*. Organisation for Economic Co-Operation and Development (OECD).

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., ... & Oskarsson, S. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, *533*(7604), 539.

Okun, A. M. (2015). *Equality and efficiency: The big tradeoff*. Brookings Institution Press.

Omoeva, C., Muoussa, W., & Hatch, R. (2018). Proposed operationalisation of equity measurement. In *Handbook of Measuring Equity in Education*. Montreal, Quebec, Canada.: UNESCO Institute for Statistics.

Ostry, J. D. (2015). Inequality and the duration of growth. *European Journal of Economics and Economic Policies: Intervention*, *12*(2), 147-157.

Ostry, M. J. D., Berg, M. A., & Tsangarides, M. C. G. (2014). *Redistribution, inequality, and growth*. International Monetary Fund.

Paccagnella, M. (2016). Age, Ageing and Skills: Results from the Survey of Adult Skills. OECD Education Working Papers, No. 132. *OECD Publishing*.

Papageorge, N. W., & Thom, K. (2018). *Genes, education, and labor market outcomes: evidence from the health and retirement study* (No. w25114). National Bureau of Economic Research.

Pareto, V. (2014). *Manual of political economy: a critical and variorum edition*. OUP Oxford.

Parfit, D. (2011a). *On what matters: volume one*. Oxford University Press.

Parfit, D. (2011b). *On what matters: volume two*. Oxford University Press.

Parfit, D. (2017). *On what matters: volume three*. Oxford University Press.

Pearl, J. (2009). *Causality*. Cambridge university press.

Pekkarinen, T., Uusitalo, R., & Kerr, S. (2009). School tracking and intergenerational income mobility: Evidence from the Finnish comprehensive school reform. *Journal of Public Economics*, *93*(7-8), 965-973.

Pekkala Kerr, S., Pekkarinen, T., & Uusitalo, R. (2013). School tracking and development of cognitive skills. *Journal of Labor Economics*, *31*(3), 577-602.

Peragine, V. (2002). Opportunity egalitarianism and income inequality. *Mathematical Social Sciences*, *44*(1), 45-64.

Pfeffer, F. T. (2008). Persistent inequality in educational attainment and its institutional context. *European sociological review*, *24*(5), 543-565.

Pfeffer, F. T. (2015). Equality and quality in education. A comparative study of 19 countries. *Social science research*, *51*, 350-368.

Pfeffer, F. T., & Hertel, F. R. (2015). How has educational expansion shaped social mobility trends in the United States?. *Social Forces*, *94*(1), 143-180.

Pietschnig, J., & Voracek, M. (2015). One century of global IQ gains: A formal meta-analysis of the Flynn effect (1909–2013). *Perspectives on Psychological Science*, *10*(3), 282-306.

Piketty, T. (2019). *Capital et idéologie*. Éditions du Seuil.

Piñeiro, G., Perelman, S., Guerschman, J. P., & Paruelo, J. M. (2008). How to evaluate models: observed vs. predicted or predicted vs. observed?. *Ecological Modelling*, *216*(3-4), 316-322.

Pinker, S. (2004). *The Blank Slate: The modern denial of human nature*. Penguin.

Plomin, R. (2018). Blueprint: How DNA makes us who we are. London, United Kingdom: Lane.

Plomin, R., & Deary, I. J. (2015). Genetics and intelligence differences: five special findings. *Molecular psychiatry*, *20*(1), 98.

Plomin, R., DeFries, J. C., & Loehlin, J. C. (1977). Genotype-environment interaction and correlation in the analysis of human behavior. *Psychological bulletin*, *84*(2), 309.

Plomin, R., & von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews Genetics*, *19*(3), 148-159.

Plug, E. J. (2001). Season of birth, schooling and earnings. *Journal of Economic Psychology*, *22*(5), 641-660.

Polderman, T. J., Benyamin, B., De Leeuw, C. A., Sullivan, P. F., Van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature genetics*, *47*(7), 702-709.

Ramos, X., & Van de Gaer, D. (2016). Approaches to inequality of opportunity: Principles, measures and evidence. *Journal of Economic Surveys*, *30*(5), 855-883.

Rawls, J. (2009). *A theory of justice*. Harvard university press.

Reardon, S. F. (2018). The Widening Academic Achievement Gap Between the Rich and the Poor. In *Inequality in the 21st Century* (pp. 177-189). Routledge.

Reich, D. (2018). *Who we are and how we got here: Ancient DNA and the new science of the human past*. Oxford University Press.

Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., ... & Albrecht, E. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *science*, *340*(6139), 1467-1471.

Rindermann, H., & Becker, D. (2018). FLynn-effect and economic growth: Do national increases in intelligence lead to increases in GDP?. *Intelligence*, *69*, 87-93.

Rindermann, H., Becker, D., & Coyle, T. R. (2017). Survey of expert opinion on intelligence: The FLynn effect and the future of intelligence. *Personality and Individual Differences*, *106*, 242-247.

Rindermann, H., & Thompson, J. (2013). Ability rise in NAEP and narrowing ethnic gaps?. *Intelligence*, *41*(6), 821-831.

Rippon, G. (2019). *The Gendered Brain: The new neuroscience that shatters the myth of the female brain*. Random House.

Ritchie, S. (2015). *Intelligence: All that matters*. Hodder & Stoughton.

Ritchie, S. J., & Bates, T. C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychological science*, *24*(7), 1301-1308.

Ritchie, S. J., Bates, T. C., & Plomin, R. (2015). Does learning to read improve intelligence? A longitudinal multivariate analysis in identical twins from age 7 to 16. *Child development*, *86*(1), 23-36.

Ritchie, S. J., & Tucker-Drob, E. M. (2018). How much does education improve intelligence? A meta-analysis. *Psychological science*, *29*(8), 1358-1369.

Roemer, J. E. (1993). A pragmatic theory of responsibility for the egalitarian planner. *Philosophy & Public Affairs*, 146-166.

Roemer, J. E. (1998). *Equality of opportunity*. Harvard University Press.

Roemer, J. E., & Elizabeth, S. (2006a). *Democracy, education, and equality: Graz-Schumpeter lectures* (No. 40). Cambridge University Press.

Roemer, J. E. (2006b). f *Economic Development as Opportunity Equalization.* Research in Economics at Yale University.

Roemer, J. E. (2012). On several approaches to equality of opportunity. *Economics & Philosophy*, *28*(2), 165-200.

Roemer, J. E. (2013). Economic development as opportunity equalization. *The World Bank Economic Review*, *28*(2), 189-209.

Roemer, J. E., & Trannoy, A. (2015). Equality of opportunity. In *Handbook of income distribution* (Vol. 2, pp. 217-300). Elsevier.

Roemer, J. E., & Trannoy, A. (2016). Equality of opportunity: Theory and measurement. *Journal of Economic Literature*, *54*(4), 1288-1332.

Roemer, J. E., & Ünveren, B. (2017). Dynamic equality of opportunity. *Economica*, *84*(334), 322-343.

Rothbard, M. N. (2000). *Egalitarianism as a revolt against nature and other essays*. Ludwig von Mises Institute.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, *66*(5), 688.

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, *100*(469), 322-331.

Rushton, J. P., & Jensen, A. R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, public policy, and law*, *11*(2), 235.

Sahlberg, P. (2007). Education policies for raising student learning: The Finnish approach. *Journal of education policy*, *22*(2), 147-171.

Sahlberg, P. (2009). A short history of educational reform in Finland. *White paper, April*. European Training Foundation ETF.

Sahlberg, P. (2014). *Finnish lessons 2.0: What can the world learn from educational change in Finland?*. Teachers College Press.

Sahlberg, P. (2011). *Finnish lessons: What can the world learn from educational change in Finland?*. Teachers College Press.

Saint-Exupéry, A. d. (1943). *Le Petit Prince*. Gallimard.

Salmela-Aro, K., & Chmielewski, A. K. (2019). Socioeconomic Inequality and Student Outcomes in Finnish Schools. In *Socioeconomic Inequality and Student Outcomes* (pp. 153-168). Springer, Singapore.

Saunders, P. (1996). *Unequal but Fair: a study of class barriers in Britain*. Institute of Economic Affairs.

Saunders, P. (2010). *Social mobility myths*. Civitas.

Scanlon, T. (1986). The significance of choice. *The tanner lectures on human values*.

Scarr-Salapatek, S. (1971). Unknowns in the IQ Equation. *Science, 174*(4015), 1223-1228. doi:10.1126/science.174.4015.1223

Schiff, M., Duyme, M., Dumaret, A., Stewart, J., Tomkiewicz, S., & Feingold, J. (1978). Intellectual status of working-class children adopted early into upper-middle-class families. *Science*, *200*(4349), 1503-1504.

Schneider, S. L. (2016). The Conceptualisation, Measurement, and Coding of Education in German and Cross-National Surveys. *GESIS Survey Guidelines*, 3. GESIS.

Schütz, G., Ursprung, H. W., & Wößmann, L. (2008). Education policy and equality of opportunity. *Kyklos*, *61*(2), 279-308.

Schwartz, S., Prins, S. J., Campbell, U. B., & Gatto, N. M. (2016). Is the" well-defined intervention assumption" politically conservative?. *Social science & medicine (1982)*, *166*, 254.

Selita, F., & Kovas, Y. (2019). Genes and Gini: what inequality means for heritability. *Journal of biosocial science*, *51*(1), 18-47.

Selzam, S., Ritchie, S. J., Pingault, J.-B., Reynolds, C. A., O'Reilly, P. F., & Plomin, R. (2019). Comparing Within- and Between-Family Polygenic Score Prediction. *The American Journal of Human Genetics*, 105(2), 351-363. doi:10.1016/j.ajhg.2019.06.006

Sen, A. K. (1992). *Inequality reexamined*. Oxford University Press.

Seppänen, P. (2006). *Kouluvalintapolitiikka perusopetuksessa: suomalaiskaupunkien koulumarkkinat kansainvälisessä valossa*. Suomen kasvatustieteellinen seura.

Seppänen, P., Kalalahti, M., Rinne, R., & Simola, H. (2015). *Lohkoutuva peruskoulu–Perheiden kouluvalinnat, yhteiskuntaluokat ja koulutuspolitiikka*. Kasvatusalan tutkimuksia, 68.

Sesardic, N. (2005). *Making sense of heritability*. Cambridge University Press.

Shayer, M., Ginsburg, D., & Coe, R. (2007). Thirty years on–a large anti-Flynn effect? The Piagetian test Volume & Heaviness norms 1975–2003. *British Journal of Educational Psychology*, *77*(1), 25-41.

Shayer, M., & Ginsburg, D. (2009). Thirty years on–a large anti-Flynn effect/(II): 13-and 14-year-olds. Piagetian tests of formal operations norms 1976–2006/7. *British Journal of Educational Psychology*, *79*(3), 409-418.

Simola, H. (2014). *The Finnish education mystery: Historical and sociological essays on schooling in Finland*. Routledge.

Simola, H., Kauko, J., Varjo, J., Kalalahti, M., & Sahlstrom, F. (2017). *Dynamics in education politics: Understanding and explaining the Finnish case*. Routledge.

Sloat, E., & Willms, J. D. (2000). The international adult literacy survey: Implications for Canadian social policy. *Canadian Journal of Education/Revue canadienne de l'education*, 218-233.

Smeeding, T., Erikson, R., & Jäntti, M. (Eds.). (2011). *Persistence, privilege, and parenting: The comparative study of intergenerational mobility*. Russell Sage Foundation.

Sohail, M., Maier, R. M., Ganna, A., Bloemendal, A., Martin, A. R., Turchin, M. C., ... & Neale, B. (2019). Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *ELife*, *8*, e39702.

Spirtes, P., Glymour, C. N., Scheines, R., & Heckerman, D. (2000). *Causation, prediction, and search*. MIT press.

Steiner, H. (2016). On the Conflict Between Liberty and Equality. In Schmidtz, D., Pavel, C. E. (Eds.) *The Oxford Handbook of Freedom*. Oxford University Press.

Steiner, P. M., Kim, Y., Hall, C. E., & Su, D. (2017). Graphical models for quasi-experimental designs. *Sociological methods & research*, *46*(2), 155-188.

Suoniemi, I. (2017). Intergenerational mobility and equal opportunity, evidence from Finland. *Labour Institute for Economic Research, Working Papers*, *312*.

te Nijenhuis, J., & Van Der Flier, H. (2013). Is the Flynn effect on g?: A meta-analysis. *Intelligence*, *41*(6), 802-807.

Teasdale, T. W., & Owen, D. R. (2008). Secular declines in cognitive test scores: A reversal of the Flynn Effect. *Intelligence*, *36*(2), 121-126.

Temkin, L. S. (1993). *Inequality*. Oxford University Press.

Temkin, L. S. (2000). Equality, Priority, and the Levelling-Down Objection. In M. Clayton & A. Williams (Eds.), *The Ideal of Equality* (pp. 126-161): Macmillan.

Temkin, L. S. (2011). Desert, Rights, Free Will, Responsibility, and Luck. *Responsibility and distributive justice*, 51.

Temkin, L. S. (2013). Equality of opportunity. In *International Encyclopedia of Ethics*. Blackwell Publishing

Temkin, L. S. (2016). The many faces of equal opportunity. *Theory and Research in Education*, *14*(3), 255-276.

Temkin, L. S. (2017). Equality as comparative fairness. *Journal of Applied Philosophy*, *34*(1), 43-60.

Trahan, L. H., Stuebing, K. K., Fletcher, J. M., & Hiscock, M. (2014). The Flynn effect: A meta-analysis. *Psychological bulletin*, *140*(5), 1332.

Tucker-Drob, E. M., & Bates, T. C. (2016). Large cross-national differences in gene× socioeconomic status interaction on intelligence. *Psychological science*, *27*(2), 138-149.

Tuddenham, R. D. (1948). Soldier intelligence in World Wars I and II. *American Psychologist*, *3*(2), 54.

Turchin, P. (2015). *Ultrasociety: How 10,000 Years of War Made Humans the Greatest Cooperators on Earth*. Beresta Books.

Turchin, P. (2017). *Ages of Discord:A Structural-Demographic Analysis of American History*. Beresta Books.

Turkheimer, E. (1998). Heritability and biological explanation. *Psychological Review*, *105*(4), 782.

Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. *Current directions in psychological science*, *9*(5), 160-164.

Turkheimer, E. (2012). Genome wide association studies of behavior are social science. In *Philosophy of behavioral biology* (pp. 43-64). Springer, Dordrecht.

Turkheimer, E. (2016). Weak genetic explanation 20 years later: reply to Plomin et al.(2016). *Perspectives on Psychological Science*, *11*(1), 24-28.

Turkheimer, E., Haley, A., Waldron, M., d'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological science*, *14*(6), 623-628.

UIS. (1999). *Operational manual for ISCED 1997 International Standard Classification of Education*. UNESCO Institute for Statistics

UIS. (2015). *ISCED 2011 Operational Manual-Guidelines for Classifying National Education Programmes and Related Qualifications*. UNESCO Institute for Statistics.

Van de Werfhorst, H. G., & Mijs, J. J. (2010). Achievement inequality and the institutional structure of educational systems: A comparative perspective. *Annual review of sociology*, *36*, 407-428.

Veldkamp, S. A., Boomsma, D. I., de Zeeuw, E. L., van Beijsterveldt, C. E., Bartels, M., Dolan, C. V., & van Bergen, E. (2019). Genetic and environmental influences on different forms of bullying perpetration, bullying victimization, and their co-occurrence. *Behavior genetics*, *49*(5), 432-443.

Willms, J. D. (1999). *Inequalities in literacy skills among youth in Canada and the United States*. Statistics Canada.

Willms, J. D. (2003). Literacy proficiency of youth: Evidence of converging socioeconomic gradients. *International Journal of Educational Research*, *39*(3), 247-252.

Wilson, R. S. (1978). Synchronies in mental development: An epigenetic perspective. *Science*, *202*(4371), 939-948.

Wilson, R. S. (1983). The Louisville twin study: Developmental synchronies in behavior. *Child development*, 298-316.

Wößmann, L. (2004). *How Equal Are Educational Opportunities? Family Background and Student Achievement in Europe and the United States* (No. 1284). Institute for the Study of Labor (IZA).

Wößmann, L. (2009). International evidence on school tracking: A review. *CESifo DICE Report*, (1), 26-34.

Wongupparaj, P., Kumari, V., & Morris, R. G. (2015). A Cross-Temporal Meta-Analysis of Raven's Progressive Matrices: Age groups and developing versus developed countries. *Intelligence*, *49*, 1-9.

Woodley, M. A. (2012). The social and scientific temporal correlates of genotypic intelligence and the Flynn effect. *Intelligence*, *40*(2), 189-204.

Woodley, M. A., & Meisenberg, G. (2013). In the Netherlands the anti-Flynn effect is a Jensen effect. *Personality and Individual Differences*, *54*(8), 871-876.

Wray, N. R., Kemper, K. E., Hayes, B. J., Goddard, M. E., & Visscher, P. M. (2019). Complex Trait Prediction from Genome Data: Contrasting EBV in Livestock to PRS in Humans: Genomic Prediction. *Genetics*, *211*(4), 1131-1141.

Young, A. I., Benonisdottir, S., Przeworski, M., & Kong, A. (2019). Deconstructing the sources of genotype-phenotype associations in humans. *Science*, *365*(6460), 1396-1400.

Young, A. I., Frigge, M. L., Gudbjartsson, D. F., Thorleifsson, G., Bjornsdottir, G., Sulem, P., ... & Kong, A. (2018). Relatedness disequilibrium regression estimates heritability without environmental bias. *Nature genetics*, *50*(9), 1304.

Young, M. (2017). *The rise of the meritocracy*. Routledge.